arXiv:1211.2522v1 [math.ST] 12 Nov 2012

The Annals of Statistics
2010, Vol. 38, No. 6, 3352–3386
DOI: 10.1214/10-AOS819
© Institute of Mathematical Statistics, 2010

IDENTIFYING THE FINITE DIMENSIONALITY OF CURVE TIME SERIES[1]

By Neil Bathia, Qiwei Yao and Flavio Ziegelmann

London School of Economics, London School of Economics and Federal University of Rio Grande do Sul

The curve time series framework provides a convenient vehicle to accommodate some nonstationary features into a stationary setup. We propose a new method to identify the dimensionality of curve time series based on the dynamical dependence across different curves. The practical implementation of our method boils down to an eigenanalysis of a finite-dimensional matrix. Furthermore, the determination of the dimensionality is equivalent to the identification of the nonzero eigenvalues of the matrix, which we carry out in terms of some bootstrap tests. Asymptotic properties of the proposed method are investigated. In particular, our estimators for zero-eigenvalues enjoy the fast convergence rate n while the estimators for nonzero eigenvalues converge at the standard $\sqrt{n}$-rate. The proposed methodology is illustrated with both simulated and real data sets.

1. Introduction. A curve time series may consist of, for example, annual weather record charts, annual production charts or daily volatility curves (from morning to evening). In these examples, the curves are segments of a single long time series. One advantage to view them as a curve series is to accommodate some nonstationary features (such as seasonal cycles or diurnal volatility patterns) into a stationary framework in a Hilbert space. There are other types of curve series that cannot be pieced together into a single long time series; for example, daily mean-variance efficient frontiers of portfolios, yield curves and intraday asset return distributions. See also an example of daily return density curves in Section 4.2. The goal of this paper is to identify the finite dimensionality of curve time series in the sense that

Received October 2009; revised February 2010.
[1]Supported in part by the Engineering and Physical Sciences Research Council of the UK.
AMS 2000 subject classifications. Primary 62M10, 62H30; secondary 60G99.
Key words and phrases. Autocovariance, curve time series, dimension reduction, eigenanalysis, Karhunen–Loéve expansion, n convergence rate, root-n convergence rate.

This is an electronic reprint of the original article published by the Institute of Mathematical Statistics in *The Annals of Statistics*, 2010, Vol. 38, No. 6, 3352–3386. This reprint differs from the original in pagination and typographic detail.

the serial dependence across different curves is driven by a finite number of scalar components. Therefore, the problem of modeling curve dynamics is reduced to that of modeling a finite-dimensional vector time series.

Throughout this paper, we assume that the observed curve time series, which we denote by $Y_1(\cdot), \ldots, Y_n(\cdot)$, are defined on a compact interval $\mathcal{I}$ and are subject to errors in the sense that

$$Y_t(u) = X_t(u) + \varepsilon_t(u), \qquad u \in \mathcal{I}, \tag{1.1}$$

where $X_t(\cdot)$ is the curve process of interest. The existence of the noise term $\varepsilon_t(\cdot)$ reflects the fact that curves $X_t(\cdot)$ are seldom perfectly observed. They are often only recorded on discrete grids and are subject to both experimental error and numerical rounding. These noisy discrete data are smoothed to yield "observed" curves $Y_t(\cdot)$. Note that both $X_t(\cdot)$ and $\varepsilon_t(\cdot)$ are unobservable.

We assume that $\varepsilon_t(\cdot)$ is a white noise sequence in the sense that $E\{\varepsilon_t(u)\} = 0$ for all t and $\mathrm{Cov}\{\varepsilon_t(u), \varepsilon_s(v)\} = 0$ for any $u, v \in \mathcal{I}$ provided $t \neq s$. This is guaranteed since we may include all the dynamic elements of $Y_t(\cdot)$ into $X_t(\cdot)$. Likewise, we may also assume that no parts of $X_t(\cdot)$ are white noise since these parts should be absorbed into $\varepsilon_t(\cdot)$. We also assume that

$$\int_{\mathcal{I}} E\{X_t(u)^2 + \varepsilon_t(u)^2\}\, du < \infty, \tag{1.2}$$

and both

$$\mu(u) \equiv E\{X_t(u)\}, \qquad M_k(u, v) \equiv \mathrm{Cov}\{X_t(u), X_{t+k}(v)\} \tag{1.3}$$

do not depend on t. Furthermore, we assume that $X_t(\cdot)$ and $\varepsilon_{t+k}(\cdot)$ are uncorrelated for all integer k. Under condition (1.2), $X_t(\cdot)$ admits the Karhunen–Loéve expansion

$$X_t(u) - \mu(u) = \sum_{j=1}^{\infty} \xi_{tj} \varphi_j(u), \tag{1.4}$$

where $\xi_{tj} = \int_{\mathcal{I}} \{X_t(u) - \mu(u)\} \varphi_j(u)\, du$ with $\{\xi_{tj}, j \geq 1\}$ being a sequence of scalar random variables with $E(\xi_{tj}) = 0$, $\mathrm{Var}(\xi_{tj}) = \lambda_j$ and $\mathrm{Cov}(\xi_{ti}, \xi_{tj}) = 0$ if $i \neq j$. We rank $\{\xi_{tj}, j \geq 1\}$ such that λ_j is monotonically decreasing as j increases.

We say that $X_t(\cdot)$ is d-dimensional if $\lambda_d \neq 0$ and $\lambda_{d+1} = 0$, where $d \geq 1$ is a finite integer; see Hall and Vial (2006). The primary goal of this paper is to identify d and to estimate the dynamic space $\mathcal{M}$ spanned by the (deterministic) eigenfunctions $\varphi_1(\cdot), \ldots, \varphi_d(\cdot)$.

Hall and Vial (2006) tackle this problem under the assumption that the curves $Y_1(\cdot), \ldots, Y_n(\cdot)$ are independent. Then the problem is insoluble in the sense that one cannot separate $X_t(\cdot)$ from $\varepsilon_t(\cdot)$ in (1.1). This difficulty was

resolved in Hall and Vial (2006) under a "low noise" setting which assumes that the noise $\varepsilon_t(\cdot)$ goes to 0 as the sample size goes to infinity. Our approach is different and it does not require the "low noise" condition, since we identify d and $\mathcal{M}$ in terms of the serial dependence of the curves. Our method relies on a simple fact that $M_k(u,v) = \mathrm{Cov}\{Y_t(u), Y_{t+k}(v)\}$ for any $k \neq 0$, which automatically filters out the noise $\varepsilon_t(\cdot)$; see (1.3). In this sense, the existence of dynamic dependence across different curves makes the problem tractable.

Dimension reduction plays an important role in functional data analysis. The most frequently used method is the functional principal component analysis in the form of applying the Karhunen–Loéve decomposition directly to the observed curves. The literature in this field is vast and includes Besse and Ramsay (1986), Dauxois, Pousse and Romain (1982), Ramsay and Dalzell (1991), Rice and Silverman (1991) and Ramsay and Silverman (2005). In spite of the methodological advancements with independent observations, the work on functional time series has been of a more theoretical nature; see, for example, Bosq (2000). The available inference methods focus mostly on nonparametric estimation for some characteristics of functional series [Part IV of Ferraty and Vieu (2006)]. As far as we are aware, the work presented here represents the first attempt on the dimension reduction based on dynamic dependence, which is radically different from the existing methods. Heuristically, our approach differs from functional principal components analysis in one fundamental manner; in principal component analysis the objective is to find the linear combinations of the data which maximize variance. In contrast, we seek for the linear combinations of the data which represent the serial dependence in the data. Although we confine ourselves to square integrable curve series in this paper, the methodology may be extended to a more general functional framework including, for example, a surface series which is particularly important for environmental study; see, for example, Guillas and Lai (2010). A follow-up study in this direction will be reported elsewhere.

The rest of the paper is organized as follows. Section 2 introduces the proposed new methodology for identifying the finite-dimensional dynamic structure. Although the Karhunen–Loéve decomposition (1.4) serves as a starting point, we do not seek for such a decomposition explicitly. Instead the eigenanalysis is performed on a positive-definite operator defined based on the autocovariance function of the curve process. Furthermore, computationally our method boils down to an eigenanalysis of a finite matrix thus requiring no computing of eigenfunctions in a functional space directly. The relevant theoretical results are presented in Section 3. As our estimation for the eigenvalues are essentially quadratic, the convergence rate of the estimators for the zero-eigenvalues is n while that for the nonzero eigenvalues is standard $\sqrt{n}$. Numerical illustration using both simulated and real datasets

is provided in Section 4. Given the nature of the subject concerned, it is inevitable to make use of some operator theory in a Hilbert space. We collect some relevant facts in Appendix A. We relegate all the technical proofs to Appendix B.

2. Methodology.

2.1. *Characterize d and $\mathcal{M}$ via serial dependence.* Let $\mathcal{L}_2(\mathcal{I})$ denote the Hilbert space consisting of all the square integrable curves defined on $\mathcal{I}$ equipped with the inner product

$$\langle f, g\rangle = \int_{\mathcal{I}} f(u)g(u)\,du, \qquad f, g \in \mathcal{L}_2(\mathcal{I}). \tag{2.1}$$

Now M_k defined in (1.3) may be viewed as the kernel of a linear operator acting on $\mathcal{L}_2(\mathcal{I})$, that is, for any $g \in \mathcal{L}_2(\mathcal{I})$, M_k maps $g(u)$ to $\breve{g}(u) \equiv \int_{\mathcal{I}} M_k(u, v)g(v)\,dv$. For notational economy, we will use M_k to denote both the kernel and the operator. Appendix A lists some relevant facts about operators in Hilbert spaces.

For M_0 defined in (1.3), we have a spectral decomposition of the form

$$M_0(u, v) = \sum_{j=1}^{\infty} \lambda_j \varphi_j(u)\varphi_j(v), \qquad u, v \in \mathcal{I}, \tag{2.2}$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$ are the eigenvalues and $\varphi_1, \varphi_2, \ldots$ are the corresponding orthonormal eigenfunctions (i.e., $\langle \varphi_i, \varphi_j\rangle = 1$ for $i = j$, and 0 otherwise). Hence,

$$\int_{\mathcal{I}} M_0(u, v)\varphi_j(v)\,dv = \lambda_j \varphi_j(u), \qquad j \geq 1.$$

Furthermore, the random curves $X_t(\cdot)$ admit the representation (1.4). We assume in this paper that $X_t(\cdot)$ is d-dimensional (i.e., $\lambda_{d+1} = 0$). Therefore,

$$M_0(u, v) = \sum_{j=1}^{d} \lambda_j \varphi_j(u)\varphi_j(v), \qquad X_t(u) = \mu(u) + \sum_{j=1}^{d} \xi_{tj}\varphi_j(u). \tag{2.3}$$

It follows from (1.1) that

$$Y_t(u) = \mu(u) + \sum_{j=1}^{d} \xi_{tj}\varphi_j(u) + \varepsilon_t(u). \tag{2.4}$$

Thus, the serial dependence of $Y_t(\cdot)$ is determined entirely by that of the d-vector process $\boldsymbol{\xi}_t \equiv (\xi_{t1}, \ldots, \xi_{td})'$ since $\varepsilon_t(\cdot)$ is white noise. By the virtue of the Karhunen–Loéve decomposition, $E\boldsymbol{\xi}_t = 0$ and $\mathrm{Var}(\boldsymbol{\xi}_t) = \mathrm{diag}(\lambda_1, \ldots, \lambda_d)$.

For some prescribed integer p, let

$$\widehat{M}_k(u,v) = \frac{1}{n-p}\sum_{j=1}^{n-p}\{Y_j(u)-\bar{Y}(u)\}\{Y_{j+k}(v)-\bar{Y}(v)\}, \tag{2.5}$$

where $\bar{Y}(\cdot) = n^{-1}\sum_{1\le j\le n} Y_j(\cdot)$ and $k = 1,\dots,p$. The reason for truncating the sums in (2.5) at $n-p$ as opposed to $n-k$ is to ensure a duality operation which simplifies the computation for eigenfunctions; see Remark 2 at the end of Section 2.2.2. The conventional approach to estimate d and $\mathcal{M} = \operatorname{span}\{\varphi_1(\cdot),\dots,\varphi_d(\cdot)\}$ is to perform an eigenanalysis on $\widehat{M}_0$ and let $\widehat{d}$ be the number of nonzero eigenvalues and $\widehat{\mathcal{M}}$ be spanned by the $\widehat{d}$ corresponding eigenfunctions; see, for example, Ramsay and Silverman (2005) and references therein. However, this approach suffers from complications due to fact that $\widehat{M}_0$ is not a consistent estimator for M_0, as $\operatorname{Cov}\{Y_t(u),Y_t(v)\} = M_0(u,v)+\operatorname{Cov}\{\varepsilon_t(u),\varepsilon_t(v)\}$. Therefore, $\widehat{M}_0$ needs to be adjusted to remove the part due to $\varepsilon_t(\cdot)$ before the eigenanalysis may be performed. Unfortunately, this is a nontrivial matter since both $X_t(\cdot)$ and $\varepsilon_t(\cdot)$ are unobservable. An alternative is to let the variance of $\varepsilon_t(\cdot)$ decay to 0 as the sample size n goes to infinity; see Hall and Vial (2006).

We adopt a different approach based on the fact that $\operatorname{Cov}\{Y_t(u),Y_{t+k}(v)\} = M_k(u,v)$ for any $k\neq 0$, which ensures that $\widehat{M}_k$ is a legitimate estimator for M_k; see (1.3) and (2.5).

Let $\mathbf{\Sigma}_k = E(\boldsymbol{\xi}_t\boldsymbol{\xi}'_{t+k}) \equiv (\sigma_{ij}^{(k)})$ be the autocovariance matrix of $\boldsymbol{\xi}_t$ at lag k. It is easy to see from (1.3) and (2.3) that $M_k(u,v) = \sum_{i,j=1}^d \sigma_{ij}^{(k)}\varphi_i(u)\varphi_j(v)$. Define a nonnegative operator

$$N_k(u,v) = \int_{\mathcal{I}} M_k(u,z)M_k(v,z)\,dz = \sum_{i,j=1}^{d} w_{ij}^{(k)}\varphi_i(u)\varphi_j(v), \tag{2.6}$$

where $\mathbf{W}_k = (w_{ij}^{(k)}) = \mathbf{\Sigma}_k\mathbf{\Sigma}'_k$ is a nonnegative definite matrix. Then it holds for any integer k that

$$\int_{\mathcal{I}} N_k(u,v)\zeta(v)\,dv = 0 \qquad \text{for any } \zeta(\cdot)\in\mathcal{M}^{\perp}, \tag{2.7}$$

where $\mathcal{M}^{\perp}$ denotes the orthogonal complement of $\mathcal{M}$ in $\mathcal{L}_2(\mathcal{I})$. Note (2.7) also holds if we replace N_k by the operator

$$K(u,v) = \sum_{k=1}^{p} N_k(u,v), \tag{2.8}$$

which is also a nonnegative operator on $\mathcal{L}_2(\mathcal{I})$.

PROPOSITION 1. *Let the matrix $\mathbf{\Sigma}_{k_0}$ be full-ranked for some $k_0 \geq 1$. Then the assertions below hold.*

(i) *The operator N_{k_0} has exactly d nonzero eigenvalues, and $\mathcal{M}$ is the linear space spanned by the corresponding d eigenfunctions.*

(ii) *For $p \geq k_0$,* (i) *also holds for the operator K.*

REMARK 1. (i) The condition that $\operatorname{rank}(\mathbf{\Sigma}_k) = d$ for some $k \geq 1$ is implied by the assumption that $X_t(\cdot)$ is d-dimensional. In the case where $\operatorname{rank}(\mathbf{\Sigma}_k) < d$ for all k, the component with no serial correlations in $X_t(\cdot)$ should be absorbed into white noise $\varepsilon_t(\cdot)$; see similar arguments on modeling vector time series in Peña and Box (1987) and Pan and Yao (2008).

(ii) The introduction of the operator K in (2.8) is to pull together the information at different lags. Using single N_k may lead to spurious choices of $\widehat{d}$.

(iii) Note that $\int_{\mathcal{I}} K(u,v)\zeta(v)\,dv = 0$ if and only if $\int_{\mathcal{I}} N_k(u,v)\zeta(v)\,dv = 0$ for all $1 \leq k \leq p$. However, we cannot use M_k directly in defining K since it does not necessarily hold that $\int_{\mathcal{I}} \sum_{1\leq k\leq p} M_k(u,v)g(v) \neq 0$ for all $g \in \mathcal{M}$. This is due to the fact that M_k are not nonnegative definite operators.

2.2. *Estimation of d and $\mathcal{M}$.*

2.2.1. *Estimators and fitted dynamic models.* Let $\psi_1, \ldots, \psi_d$ be the orthonormal eigenfunctions of K corresponding to its d nonzero eigenvalues. Then they form an orthonormal basis of $\mathcal{M}$; see Proposition 1(ii) above. Hence, it holds that

$$X_t(u) - \mu(u) = \sum_{j=1}^{d} \xi_{tj}\varphi_j(u) = \sum_{j=1}^{d} \eta_{tj}\psi_j(u),$$

where $\eta_{tj} = \int_{\mathcal{I}} \{X_t(u) - \mu(u)\}\psi_j(u)\,du$. Therefore, the serial dependence of $X_t(\cdot)$ [and also that of $Y_t(\cdot)$] can be represented by that of the d-vector process $\boldsymbol{\eta}_t \equiv (\eta_{t1}, \ldots, \eta_{td})'$. Since (ξ_{tj}, φ_j) cannot be estimated directly from Y_t (see Section 2.1 above), we estimate (η_{tj}, ψ_j) instead.

As we have stated above, M_k for $k \neq 0$ may be directly estimated from the observed curves Y_t; see (2.5). Hence, a natural estimator for K may be defined as

$$\begin{aligned}\widehat{K}(u,v) &= \sum_{k=1}^{p} \int_{\mathcal{I}} \widehat{M}_k(u,z)\widehat{M}_k(v,z)\,dz \\ &= \frac{1}{(n-p)^2} \sum_{t,s=1}^{n-p} \sum_{k=1}^{p} \{Y_t(u) - \bar{Y}(u)\}\{Y_s(v) - \bar{Y}(v)\} \\ &\qquad \times \langle Y_{t+k} - \bar{Y}, Y_{s+k} - \bar{Y}\rangle, \end{aligned} \tag{2.9}$$

see (2.8), (2.6), (2.5) and (2.1).

By Proposition 1, we define $\widehat{d}$ to be the number of nonzero eigenvalues of $\widehat{K}$ (see Section 2.2.3 below) and $\widehat{\mathcal{M}}$ to be the linear space spanned by the $\widehat{d}$ corresponding orthonormal eigenfunctions $\widehat{\psi}_1(\cdot), \ldots, \widehat{\psi}_{\widehat{d}}(\cdot)$. This leads to the fitting

$$\widehat{Y}_t(u) = \bar{Y}(u) + \sum_{j=1}^{\widehat{d}} \widehat{\eta}_{tj} \widehat{\psi}_j(u), \qquad u \in \mathcal{I}, \tag{2.10}$$

where

$$\widehat{\eta}_{tj} = \int_{\mathcal{I}} \{Y_t(u) - \bar{Y}(u)\} \widehat{\psi}_j(u)\, du, \qquad j = 1, \ldots, \widehat{d}. \tag{2.11}$$

Although $\widehat{\psi}_j$ are not the estimators for the eigenfunctions φ_j of M_0 defined in (2.2), $\widehat{\mathcal{M}} = \operatorname{span}\{\widehat{\psi}_1(\cdot), \ldots, \widehat{\psi}_d(\cdot)\}$ is a consistent estimator of $\mathcal{M} = \operatorname{span}\{\varphi_1(\cdot), \ldots, \varphi_d(\cdot)\}$ (Theorem 2 in Section 3 below).

In order to model the dynamic behavior of $Y_t(\cdot)$, we only need to model the $\widehat{d}$-dimensional vector process $\widehat{\boldsymbol{\eta}}_t \equiv (\widehat{\eta}_{t1}, \ldots, \widehat{\eta}_{t\widehat{d}})'$; see (2.10) above. This may be done using VARMA or any other multivariate time series models. See also Tiao and Tsay (1989) for applying linear transformations in order to obtain a more parsimonious model for $\widehat{\boldsymbol{\eta}}_t$.

The integer p used in (2.5) may be selected in the same spirit as the maximum lag used in, for example, the Ljung–Box–Pierce portmanteau test for white noise. In practice, we often choose p to be a small positive integer. Note that k_0 fulfilling the condition of Proposition 1 is often small since serial dependence decays as the lag increases for most practical data.

2.2.2. *Eigenanalysis.* To perform an eigenanalysis in a Hilbert space is not a trivial matter. A popular pragmatic approach is to use an approximation via discretization, that is, to evaluate the observed curves at a fine grid and to replace the observed curves by the resulting vectors. This is an approximate method; effectively transform the problem to an eigenanalysis for a finite matrix. See, for example, Section 8.4 of Ramsay and Silverman (2005). Below we also transform the problem into an eigenanalysis of a finite matrix but not via any approximations. Instead we make use of the well-known duality property that $\mathbf{AB}'$ and $\mathbf{B}'\mathbf{A}$ share the same nonzero eigenvalues for any matrices $\mathbf{A}$ and $\mathbf{B}$ of the same sizes. Furthermore, if $\boldsymbol{\gamma}$ is an eigenvector of $\mathbf{B}'\mathbf{A}$, $\mathbf{A}\boldsymbol{\gamma}$ is an eigenvector of $\mathbf{AB}'$ with the same eigenvalue. In fact, this duality also holds for operators in a Hilbert space. This scheme was adopted in Kneip and Utikal (2001) and Benko, Hardle and Kneip (2009).

We present a heuristic argument first. To view the operator $\widehat{K}(\cdot, \cdot)$ defined in (2.9) in the form of $\mathbf{AB}'$, let us denote the curve $Y_t(\cdot) - \bar{Y}(\cdot)$ as an $\infty \times 1$ vector $\mathbf{Y}_t$ with $\mathbf{Y}_t'\mathbf{Y}_s = \langle Y_t - \bar{Y}, Y_s - \bar{Y} \rangle$; see (2.1). Put $\mathcal{Y}_k =$

$(\mathbf{Y}_{1+k},\dots,\mathbf{Y}_{n-p+k})$. Then $\widehat{K}(\cdot,\cdot)$ may be represented as an $\infty\times\infty$ matrix

$$\widehat{\mathbf{K}}=\frac{1}{(n-p)^2}\mathcal{Y}_0\sum_{k=1}^{p}\mathcal{Y}_k'\mathcal{Y}_k\mathcal{Y}_0'.$$

Applying the duality with $\mathbf{A}=\mathcal{Y}_0$ and $\mathbf{B}'=\sum_{1\le k\le p}\mathcal{Y}_k'\mathcal{Y}_k\mathcal{Y}_0'$, $\widehat{\mathbf{K}}$ shares the same nonzero eigenvalues with the $(n-p)\times(n-p)$ matrix

$$\mathbf{K}^*=\frac{1}{(n-p)^2}\sum_{k=1}^{p}\mathcal{Y}_k'\mathcal{Y}_k\mathcal{Y}_0'\mathcal{Y}_0, \tag{2.12}$$

where the (t,s)th element of $\mathcal{Y}_k'\mathcal{Y}_k$ is $\mathbf{Y}_{t+k}'\mathbf{Y}_{s+k}=\langle Y_{t+k}-\bar{Y},Y_{s+k}-\bar{Y}\rangle$ and $k=0,1,\dots,p$. Furthermore, let $\boldsymbol{\gamma}_j=(\gamma_{1j},\dots,\gamma_{n-p,j})'$, $j=1,\dots,\widehat{d}$, be the eigenvectors of $\mathbf{K}^*$ corresponding to the $\widehat{d}$ largest eigenvalues. Then

$$\sum_{t=1}^{n-p}\gamma_{tj}\{Y_t(\cdot)-\bar{Y}(\cdot)\},\qquad j=1,\dots,\widehat{d}, \tag{2.13}$$

are the $\widehat{d}$ eigenfunctions of $\widehat{K}(\cdot,\cdot)$. Note that the functions in (2.13) may not be orthogonal with each other. Thus, the orthonormal eigenfunctions $\widehat{\psi}_1(\cdot),\dots,\widehat{\psi}_{\widehat{d}}(\cdot)$ used in (2.10) may be obtained by applying a Gram–Schmidt algorithm to the functions given in (2.13).

The heuristic argument presented above is justified by result below. The formal proof is relegated to Appendix B.

PROPOSITION 2. *The operator $\widehat{K}(\cdot,\cdot)$ shares the same nonzero eigenvalues with matrix $\mathbf{K}^*$ defined in (2.12) with the corresponding eigenfunctions given in (2.13).*

REMARK 2. The truncation of the sums in (2.5) at $(n-p)$ for different k is necessary to ensure the applicability of the above duality operation. If we truncated the sum for $\widehat{M}_k$ at $(n-k)$ instead, $\mathcal{Y}_k'\mathcal{Y}_k$ would be of different sizes for different k, and $\mathbf{K}^*$ in (2.12) would not be well defined.

2.2.3. *Determination of d via statistical tests.* Although the number of nonzero eigenvalues of operator $K(\cdot,\cdot)$ defined in (2.8) is d [Proposition 1(ii)], the number of nonzero eigenvalues of its estimator $\widehat{K}(\cdot,\cdot)$ defined in (2.9) may be much greater than d due to random fluctuation in the sample. One empirical approach is to take $\widehat{d}$ to be the number of "large" eigenvalues of $\widehat{K}$ in the sense that the $(\widehat{d}+1)$th largest eigenvalue drops significantly; see also Theorem 3 in Section 3 and Figure 1 in Section 4.1. Hyndman and Ullah (2007) proposed to choose d by minimizing forecasting errors. Below, we present a bootstrap test to determine the value of d.

Let $\theta_1 \geq \theta_2 \geq \cdots \geq 0$ be the eigenvalues of K. If the true dimensionality is $d = d_0$, we expect to reject the null hypothesis $\theta_{d_0} = 0$, and not to reject the hypothesis $\theta_{d_0+1} = 0$. Suppose we are interested in testing the null hypothesis

$$H_0 : \theta_{d_0+1} = 0, \tag{2.14}$$

where d_0 is a known integer, obtained, for example, by visual observation of the estimated eigenvalues $\widehat{\theta}_1 \geq \widehat{\theta}_2 \geq \cdots \geq 0$ of $\widehat{K}$. Hence, we reject H_0 if $\hat{\theta}_{d_0+1} > l_\alpha$, where l_α is the critical value at the $\alpha \in (0,1)$ significance level. To evaluate the critical value l_α, we propose the following bootstrap procedure.

1. Let $\widehat{Y}_t(\cdot)$ be defined as in (2.10) with $\widehat{d} = d_0$. Let $\widehat{\varepsilon}_t(\cdot) = Y_t(\cdot) - \widehat{Y}_t(\cdot)$.
2. Generate a bootstrap sample from the model
$$Y_t^*(\cdot) = \widehat{Y}_t(\cdot) + \varepsilon_t^*(\cdot),$$
where ε_t^* are drawn independently (with replacement) from $\{\widehat{\varepsilon}_1, \ldots, \widehat{\varepsilon}_n\}$.
3. Form an operator K^* in the same manner as $\widehat{K}$ with $\{Y_t\}$ replaced by $\{Y_t^*\}$, compute the (d_0+1)th largest eigenvalue $\theta_{d_0+1}^*$ of K^*.

Then the conditional distribution of $\theta_{d_0+1}^*$, given the observations $\{Y_1, \ldots, Y_n\}$, is taken as the distribution of $\hat{\theta}_{d_0+1}$ under H_0. In practical implementation, we repeat Steps 2 and 3 above B times for some large integer B, and we reject H_0 if the event that $\theta_{d_0+1}^* > \hat{\theta}_{d_0+1}$ occurs not more than $[\alpha B]$ times. The simulation results reported in Section 4.1 below indicate that the above bootstrap method works well.

REMARK 3. The serial dependence in X_t could provide an alternative method for testing hypothesis (2.14). Under model (2.4), the projected series of the curves $Y_t(\cdot)$ on any direction perpendicular to $\mathcal{M}$ is white noise. Put $U_t = \langle Y_t, \widehat{\psi}_{d_0+1} \rangle, t = 1, \ldots, n$. Then U_t would behave like a (scalar) white noise under H_0. However, for example, the Ljung–Box–Pierce portmanteau test for white noise coupled with the standard χ^2-approximation does not work well in this context. This is due to the fact that the $(d+1)$th largest eigenvalue $\widehat{K}$ is effectively the extreme value of the estimates for all the zero-eigenvalues of K. Therefore, $\widehat{\psi}_{d_0+1}$ is not an estimate for a fixed direction, which makes the χ^2-approximation for the Ljung–Box–Pierce statistic mathematically invalid. Indeed some simulation results, not reported here, indicate that the χ^2-approximation tends to underestimate the critical values for the Ljung–Box–Pierce test in this particular context.

3. Theoretical properties. Before presenting the asymptotic results, we first solidify some notation. Denote by (θ_j, ψ_j) and $(\widehat{\theta}_j, \widehat{\psi}_j)$ the (eigenvalue, eigenfunction) pairs of K and $\widehat{K}$, respectively [see (2.8) and (2.9)]. We always arrange the eigenvalues in descending order, that is, $\theta_j > \theta_{j+1}$. As the eigenfunctions of K and $\widehat{K}$ are unique only up to sign changes, in the sequel,

it will go without saying that the right versions are used. Furthermore, recall that $\theta_j = 0$ for all $j \geq d+1$. Thus, the eigenfunctions ψ_j are not identified for $j \geq d+1$. We take this last point into consideration in our theory. We always assume that the dimension $d \geq 1$ is a fixed finite integer, and $p \geq 1$ is also a fixed finite integer.

For simplicity in the proofs, we suppose that $E\{Y_t(\cdot)\} = \mu(\cdot)$ is known and thus set $\bar{Y}(\cdot) = \mu(\cdot)$. Straightforward adjustments to our arguments can be made when this is not the case. We denote by $\|L\|_{\mathcal{S}}$ the Hilbert–Schmidt norm for any operator L; see Appendix A. Our asymptotic results are based on the following regularity conditions:

C1. $\{Y_t(\cdot)\}$ is strictly stationary and ψ-mixing with the mixing coefficient defined as

$$\psi(l) = \sup_{A \in \mathcal{F}_{\infty}^{0}, B \in \mathcal{F}_{l}^{\infty}, P(A)P(B)>0} |1 - P(B|A)/P(B)|,$$

where $\mathcal{F}_i^j = \sigma\{Y_i(\cdot), \ldots, Y_j(\cdot)\}$ for any $j \geq i$. In addition, it holds that $\sum_{l=1}^{\infty} l \times \psi^{1/2}(l) < \infty$.

C2. $E\{\int_{\mathcal{I}} Y_t(u)^2 \, du\}^2 < \infty$.

C3. $\theta_1 > \cdots > \theta_d > 0 = \theta_{d+1} = \cdots$, that is, all the nonzero eigenvalues of K are different.

C4. $\mathrm{Cov}\{X_s(u), \varepsilon_t(v)\} = 0$ for all s, t and $u, v \in \mathcal{I}$.

THEOREM 1. *Let conditions* C1–C4 *hold. Then as* $n \to \infty$, *the following assertions hold:*

(i) $\|\widehat{K} - K\|_{\mathcal{S}} = O_p(n^{-1/2})$.

(ii) *For* $j = 1, \ldots, d$, $|\widehat{\theta}_j - \theta_j| = O_p(n^{-1/2})$ *and*

$$\left(\int_{\mathcal{I}} \{\widehat{\psi}_j(u) - \psi_j(u)\}^2 \, du \right)^{1/2} = O_P(n^{-1/2}).$$

(iii) *For* $j \geq d+1$, $\widehat{\theta}_j = O_p(n^{-1})$.

(iv) *Let* $\{\psi_j : j \geq d+1\}$ *be a complete orthonormal basis of* $\mathcal{M}^{\perp}$, *and put*

$$f_j(\cdot) = \sum_{i=d+1}^{\infty} \langle \psi_i, \widehat{\psi}_j \rangle \psi_i(\cdot).$$

Then for any $j \geq d+1$,

$$\left(\int_{\mathcal{I}} \left\{ \sum_{i=1}^{d} \langle \psi_i, \widehat{\psi}_j \rangle \psi_i(u) \right\}^2 du \right)^{1/2} = \left(\int_{\mathcal{I}} \{\widehat{\psi}_j(u) - f_j(u)\}^2 \, du \right)^{1/2} = O_p(n^{-1/2}).$$

REMARK 4. (a) In the above theorem, assertions (i) and (ii) are standard. (In fact, those results still hold for $d = \infty$.)

(b) Assertion (iv) implies that the estimated eigenfunctions $\widehat{\psi}_{d+j}$, $j \geq 1$, are asymptotically in the orthogonal complement of the dynamic space $\mathcal{M}$.

(c) The fast convergence rate n in assertion (iii) deserves some further explanation. To this end, we consider a simple analogue: let $A_1, \ldots, A_n$ be a sample of stationary random variables, and we are interested in estimating $\mu^2 = (EA_t)^2$ for which we use the estimator $\bar{A}^2 = (n^{-1}\sum_{t=1}^n A_t)^2 = n^{-2}\sum_{s,t=1}^n A_s A_t$. Then under appropriate regularity conditions, it holds that

$$|\bar{A}^2 - \mu^2| \leq |\mu||\bar{A} - \mu| + |\bar{A}^2 - \bar{A}\mu| = |\mu| \cdot O_p(n^{-1/2}) + O_p(n^{-1}) \tag{3.1}$$

as $|\bar{A} - \mu| = O_p(n^{-1/2})$ and $|\bar{A}^2 - \bar{A}\mu| = O_p(n^{-1})$. The latter follows from a simple U-statistic argument; see Lee (1990). It is easy to see from (3.1) that $|\bar{A}^2 - \mu^2| = O_p(n^{-1/2})$ if $\mu \neq 0$, and $|\bar{A}^2 - \mu^2| = O_p(n^{-1})$ if $\mu = 0$. In our context, the operator $\widehat{K} = \sum_{k=1}^p \int_{\mathcal{I}} M_k(u,r)M_k(v,r) = (n-p)^{-2}\sum_{k=1}^p \sum_{s,t=1}^{n-p} Z_{ik}Z_{jk}^*(u,v)$, where $Z_{tk}(u,v) = \{Y_t(u) - \mu(u)\}\{Y_{t+k}(v) - \mu(v)\}$ and $Z_{ik}Z_{jk}^*(u,v) = \int_{\mathcal{I}} Z_{ik}(u,r)Z_{jk}(v,r)\,dr$, is similar to $\bar{A}^2$, and hence the convergence properties stated in Theorem 1(iii) [and also (ii)]. The fast convergence rate, which is termed as "super-consistent" in econometric literature, is illustrated via simulation in Section 4.1 below; see Figures 4–7. It makes the identification of zero-eigenvalues easier; see Figure 1.

With d known, let $\widetilde{\mathcal{M}} = \operatorname{span}\{\widehat{\psi}_1(\cdot), \ldots, \widehat{\psi}_d(\cdot)\}$, where $\widehat{\psi}_1(\cdot), \ldots, \widehat{\psi}_d(\cdot)$ are the eigenfunctions of $\widehat{K}$ corresponding to the d largest eigenvalues. In order to measure the discrepancy between $\mathcal{M}$ and $\widetilde{\mathcal{M}}$, we introduce the following metric. Let $\mathcal{N}_1$ and $\mathcal{N}_2$ be any two d-dimensional subspaces of $\mathcal{L}_2(\mathcal{I})$. Let $\{\zeta_{i1}(\cdot), \ldots, \zeta_{id}(\cdot)\}$ be an orthonormal basis of $\mathcal{N}_i$, $i = 1, 2$. Then the projection of ζ_{1k} onto $\mathcal{N}_2$ may be expressed as

$$\sum_{j=1}^d \langle \zeta_{2j}, \zeta_{1k} \rangle \zeta_{2j}(u).$$

Its squared norm is $\sum_{j=1}^d (\langle \zeta_{2j}, \zeta_{1k} \rangle)^2 \leq 1$. The discrepancy measure is defined as

$$D(\mathcal{N}_1, \mathcal{N}_2) = \sqrt{1 - \frac{1}{d}\sum_{j,k=1}^d (\langle \zeta_{2j}, \zeta_{1k} \rangle)^2}. \tag{3.2}$$

It is clear that this is a symmetric measure between 0 and 1. It is independent of the choice of the orthonormal bases used in the definition, and it equals 0 if and only if $\mathcal{N}_1 = \mathcal{N}_2$. Let $\mathcal{Z}$ be the set consisting of all the d-dimensional subspaces in $\mathcal{L}_2(\mathcal{I})$. Then $(\mathcal{Z}, D)$ forms a metric space in the sense that D is a well-defined distance measure on $\mathcal{Z}$ (see Lemma 4 in Appendix B below).

THEOREM 2. *Let the conditions of Theorem 1 hold. Suppose that d is known. Then as $n \to \infty$, it holds that $D(\widetilde{\mathcal{M}}, \mathcal{M}) = O_p(n^{-1/2})$.*

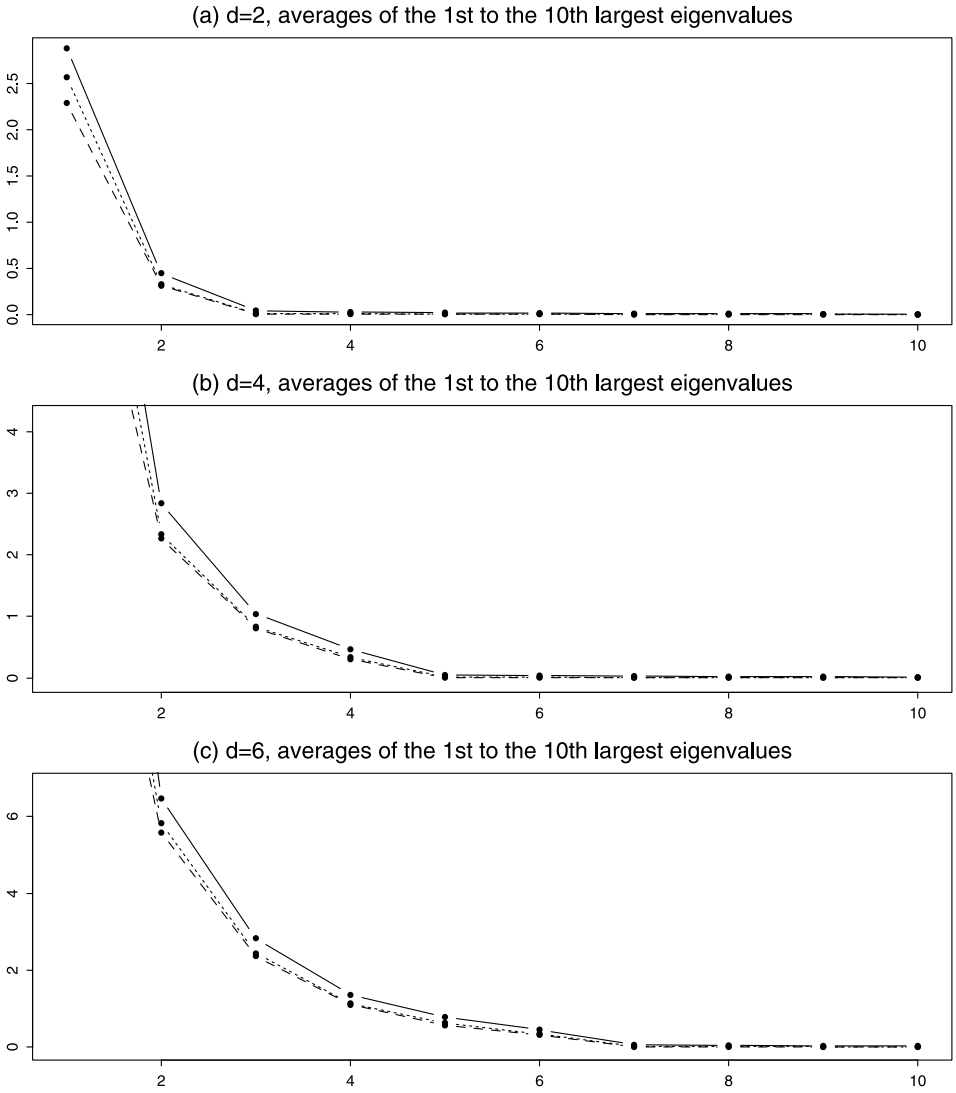

FIG. 1. *The average estimated eigenvalues over the 200 replications with sample sizes* $n = 100$ *(solid lines),* 300 *(dotted lines) and* 600 *(dashed lines).*

REMARK 5. Our estimation of $\mathcal{M}$ is asymptotically adaptive to d. To this end, let $\widehat{d}$ be a consistent estimator of d in the sense that $P(\widehat{d} = d) \to 1$, and $\widehat{\mathcal{M}} = \text{span}\{\widehat{\psi}_1, \ldots, \widehat{\psi}_{\widehat{d}}\}$ be the estimator of $\mathcal{M}$ with d estimated by $\widehat{d}$. Since $\widehat{d}$ may differ from d, we use the modified metric $\widetilde{D}$, defined in (4.1) below, to measure the difference between $\widehat{\mathcal{M}}$ and $\mathcal{M}$. Then it holds for any constant $C > 0$ that

$$P\{n^{1/2}|\widetilde{D}(\widehat{\mathcal{M}}, \mathcal{M}) - D(\widehat{\mathcal{M}}, \mathcal{M})| > C\}$$

$$\le P\{n^{1/2}|\widetilde{D}(\widehat{\mathcal{M}},\mathcal{M}) - D(\widetilde{\mathcal{M}},\mathcal{M})| > C|\widehat{d}=d\}P(\widehat{d}=d) + P(\widehat{d}\neq d)$$
$$\le P\{n^{1/2}|\widetilde{D}(\widehat{\mathcal{M}},\mathcal{M}) - D(\widetilde{\mathcal{M}},\mathcal{M})| > C|\widehat{d}=d\} + o(1).$$

Note that when $\widehat{d}=d$, $\widehat{\mathcal{M}}=\widetilde{\mathcal{M}}$ and thus $\widetilde{D}(\widehat{\mathcal{M}},\mathcal{M}) = D(\widehat{\mathcal{M}},\mathcal{M})$. Hence the conditional probability on the RHS of the above expression is 0. This together with Theorem 2 yield $\widetilde{D}(\widehat{\mathcal{M}},\mathcal{M}) = O_p(n^{-1/2})$.

One such consistent estimator of d may be defined as $\widehat{d} = \#\{j : \widehat{\theta}_j \ge \epsilon\}$, where $\epsilon = \epsilon(n) > 0$ satisfies the conditions in Theorem 3 below.

THEOREM 3. *Let the conditions of Theorem 1 hold. Let* $\epsilon \to 0$ *and* $\epsilon^2 n \to \infty$ *and as* $n \to \infty$. *Then* $P(\widehat{d} \neq d) \to 0$.

4. Numerical properties.

4.1. *Simulations.* We illustrate the proposed method first using the simulated data from model (1.1) with

$$X_t(u) = \sum_{i=1}^{d} \xi_{ti}\varphi_i(u), \qquad \varepsilon_t(u) = \sum_{j=1}^{10} \frac{Z_{tj}}{2^{j-1}}\zeta_j(u), \qquad u \in [0,1],$$

where $\{\xi_{ti}, t \ge 1\}$ is a linear AR(1) process with the coefficient $(-1)^i(0.9 - 0.5i/d)$, the innovations Z_{tj} are independent $N(0,1)$ variables and

$$\varphi_i(u) = \sqrt{2}\cos(\pi i u), \qquad \zeta_j(u) = \sqrt{2}\sin(\pi j u).$$

We set sample size $n = 100, 300$ or 600, and the dimension parameter $d = 2, 4$ or 6. For each setting, we repeat the simulation 200 times. We use $p = 5$ in defining the operator $\widehat{K}$ in (2.9). For each of the 200 samples, we replicate the bootstrap sampling 200 times.

The average of the ordered eigenvalues of $\widehat{K}$ obtained from the 200 replications are plotted in Figure 1. For a good visual illustration, we only plot the ten largest eigenvalues. It is clear that drop from the dth largest eigenvalue to the $(d+1)$st is very pronounced. Furthermore, the estimates for zero-eigenvalues with different sample size are much closer than those for nonzero eigenvalues. This evidence is in line with the different convergence rates presented in Theorem 1(ii) and (iii). We apply the bootstrap method to test the hypothesis that the dth or the $(d+1)$st largest eigenvalue of K (θ_d and θ_{d+1}, resp.) are 0. The results are summarized in Figure 2. The bootstrap test cannot reject the true null hypothesis $\theta_{d+1} = 0$. The false null hypothesis $\theta_d = 0$ is routinely rejected when $n = 600$ or 300; see Figure 2(a). However, the test does not work when the sample size is as small as 100.

To measure the accuracy of the estimation for the factor loading space $\mathcal{M}$, we need to modify the metric D defined in (3.2) first, as $\widehat{d}$ may be different from d. Let $\mathcal{N}_1, \mathcal{N}_2$ be two subspaces in $\mathcal{L}_2(\mathcal{I})$ with dimension d_1 and d_2,

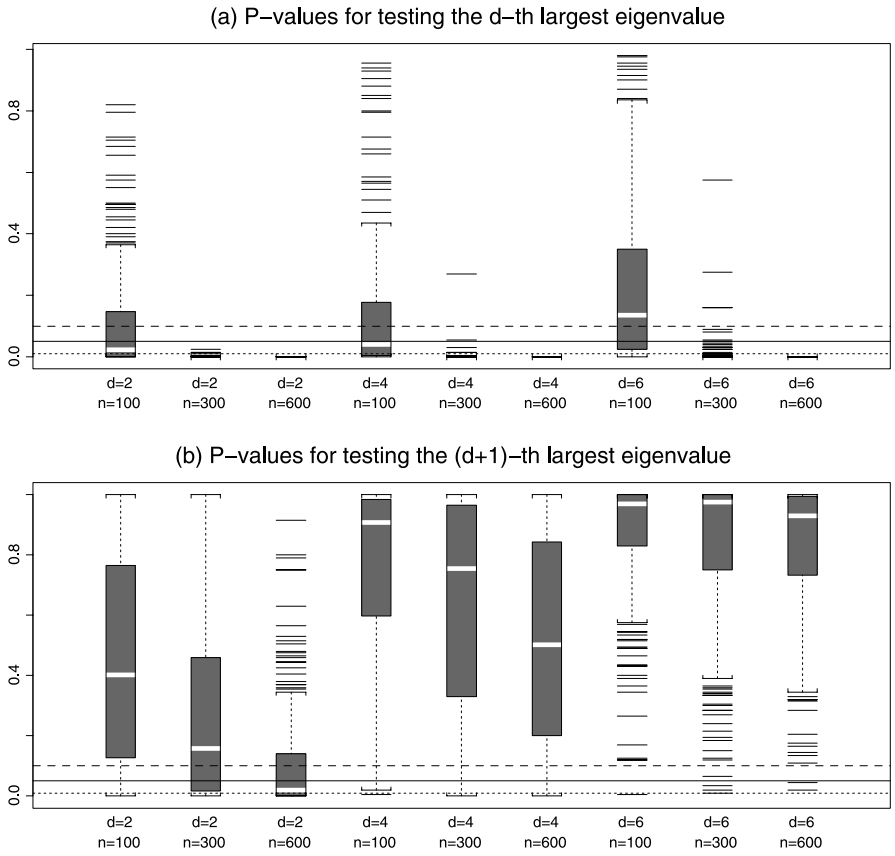

FIG. 2. *The boxplots of the P-values for the bootstrap tests of the hypothesis that* (a) *the dth largest eigenvalue of K is 0, and* (b) *the* $(d+1)$*th largest eigenvalue of K is 0. The horizontal lines mark the 1% (dotted line), 5% (solid lines) and 10% (dashed lines) significance levels, respectively.*

respectively. Let $\{\zeta_{i1}, \ldots, \zeta_{id_i}\}$ be an orthonormal basis of $\mathcal{N}_i$, $i = 1, 2$. The discrepancy measure between the two subspaces is defined as

$$\widetilde{D}(\mathcal{N}_1, \mathcal{N}_2) = \sqrt{1 - \frac{1}{\max(d_1, d_2)} \sum_{k=1}^{d_1} \sum_{j=1}^{d_2} (\langle \zeta_{2j}, \zeta_{1k} \rangle)^2}. \tag{4.1}$$

It can be shown that $\widetilde{D}(\mathcal{N}_1, \mathcal{N}_2) \in [0, 1]$. It equals 0 if and only if $\mathcal{N}_1 = \mathcal{N}_2$, and 1 if and only if $\mathcal{N}_1 \perp \mathcal{N}_2$. Obviously, $\widetilde{D}(\mathcal{N}_1, \mathcal{N}_2) = D(\mathcal{N}_1, \mathcal{N}_2)$ when $d_1 = d_2 = d$. We computed $\widetilde{D}(\widehat{\mathcal{M}}, \mathcal{M})$ in the 200 replications for each setting. Figure 3 presents the boxplots of those $\widetilde{D}$-values. It is noticeable that the $\widetilde{D}$ measure decreases as the sample size n increases. It is interesting to note too that the accuracy of the estimation is independent of the dimension d.

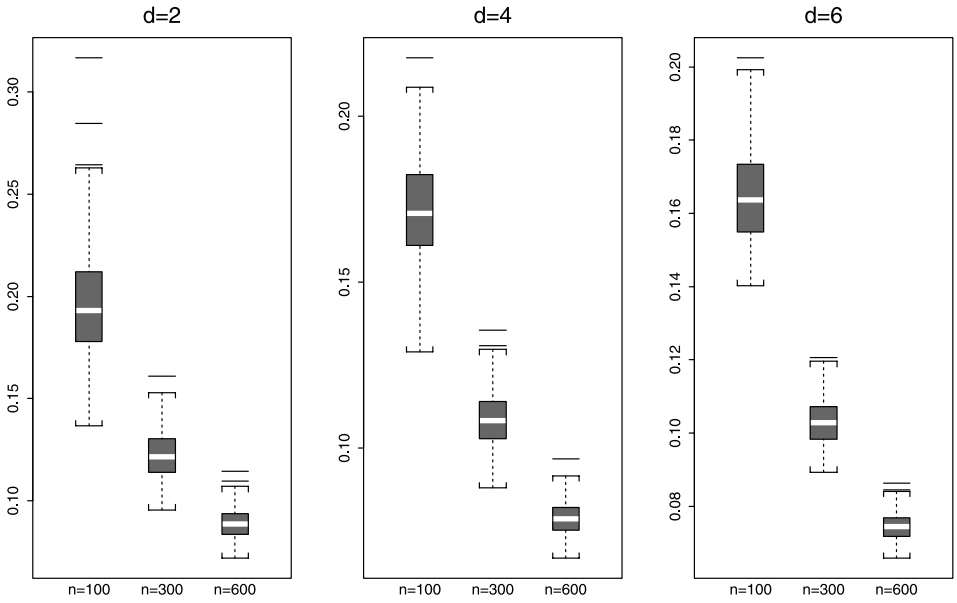

FIG. 3. *The boxplots for the estimated error $\widetilde{D}$ defined in (4.1).*

To further illustrate the different convergence rates in estimating nonzero and zero eigenvalues, as stated in Theorem 1, we generate 10,000 samples with different sample sizes from model (1.1) with $d=1$, $\xi_t = 0.5\xi_{t-1} + \eta_t$, where $\eta_t \sim N(0,1)$, $\varphi(u) = \sqrt{2}\cos(\pi u)$, and $\varepsilon_t(\cdot)$ is the same as above. In defining the operator K, we let $p=1$. Then the operator K has only one nonzero eigenvalue $\theta = 2$. Figure 4 depicts the standardized histograms and the kernel density estimators of $\sqrt{n}(\widehat{\theta}_1 - \theta)$, computed from the 10,000 samples. It is evident that those distributions resemble normal distributions when the sample size is 200 or greater. This is in line with Theorem 1(ii) which implies that $\sqrt{n}(\widehat{\theta}_1 - \theta)$ converges to a nondegenerate distribution.

Figure 5 displays the distribution of $\sqrt{n}\widehat{\theta}_2$, noting $\theta_2 = 0$. It is clear that $\sqrt{n}\widehat{\theta}_2$ converges to zero as n increases, indicating the fact that the normalized factor $\sqrt{n}$ is too small to stabilize the distribution. In contrast, Figure 6 exhibits that the distribution of $n\widehat{\theta}_2$ stabilizes from the sample size as small as $n = 50$; see Theorem 1(iii). In fact, the profile of the distribution with $n = 10$ looks almost the same as that with $n = 2000$.

Figure 7 displays boxplots of the absolute estimation errors of the eigenvalues. With the same sample size, the estimation errors for the nonzero eigenvalue are considerably greater that those for the zero eigenvalue.

4.2. *A real data example.* To further illustrate the methodology developed in this paper, we set upon the task of modeling the intraday return densities for the IBM stock in 2006. To this end, we have obtained the intra-

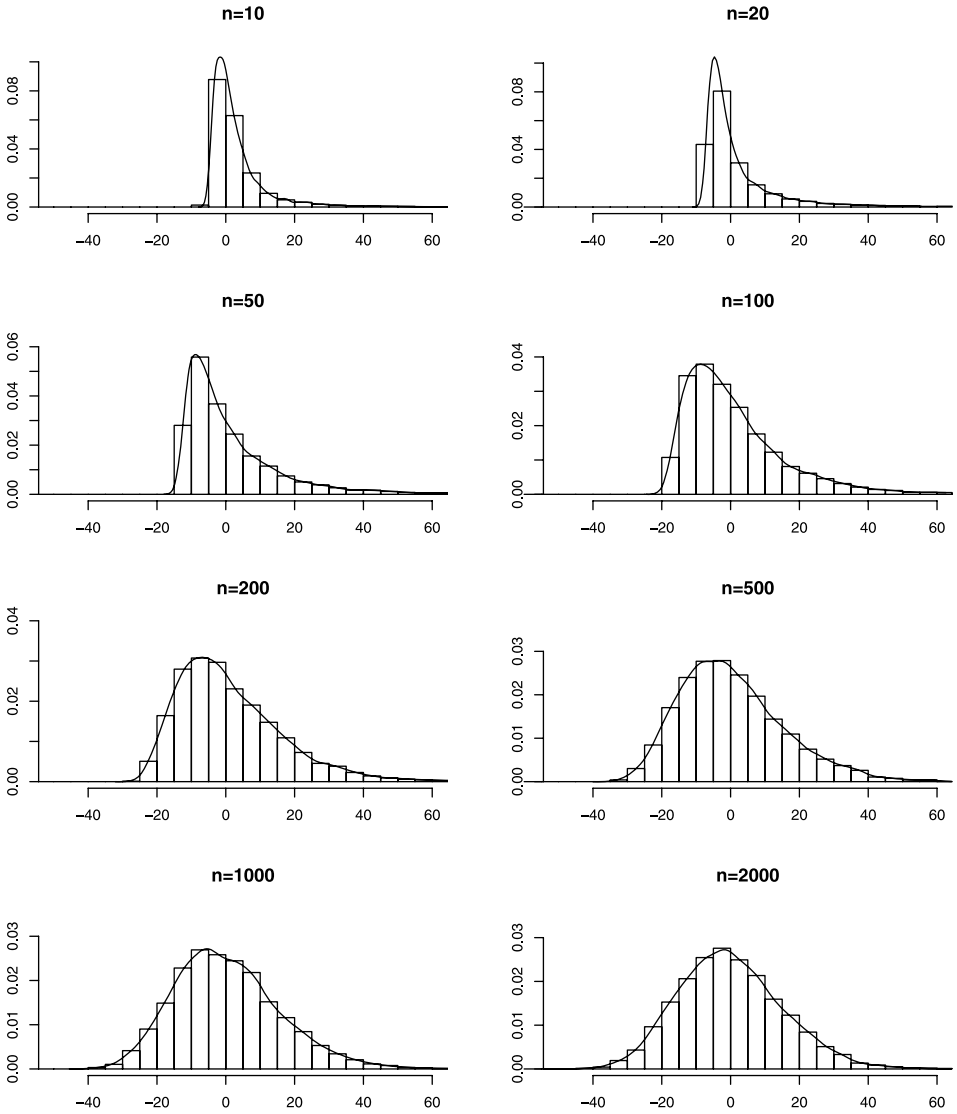

FIG. 4. *Standardized histograms overlaid by kernel density estimators of* $\sqrt{n}(\widehat{\theta}_1 - \theta)$.

day prices via the WRDS database. We only use prices between 09:30–16:00 since the market is not particularly active outside of these times. There are $n = 251$ trading days in the sample and a total of 2,786,650 observations. The size of this dataset is 73.7 MB.

Since high frequency prices are not equally spaced in time, we compute the returns using the prices at the so-called previous tick times in every 5 minute intervals. More precisely, we set the sampling times at $\tau_1 = 09{:}35$,

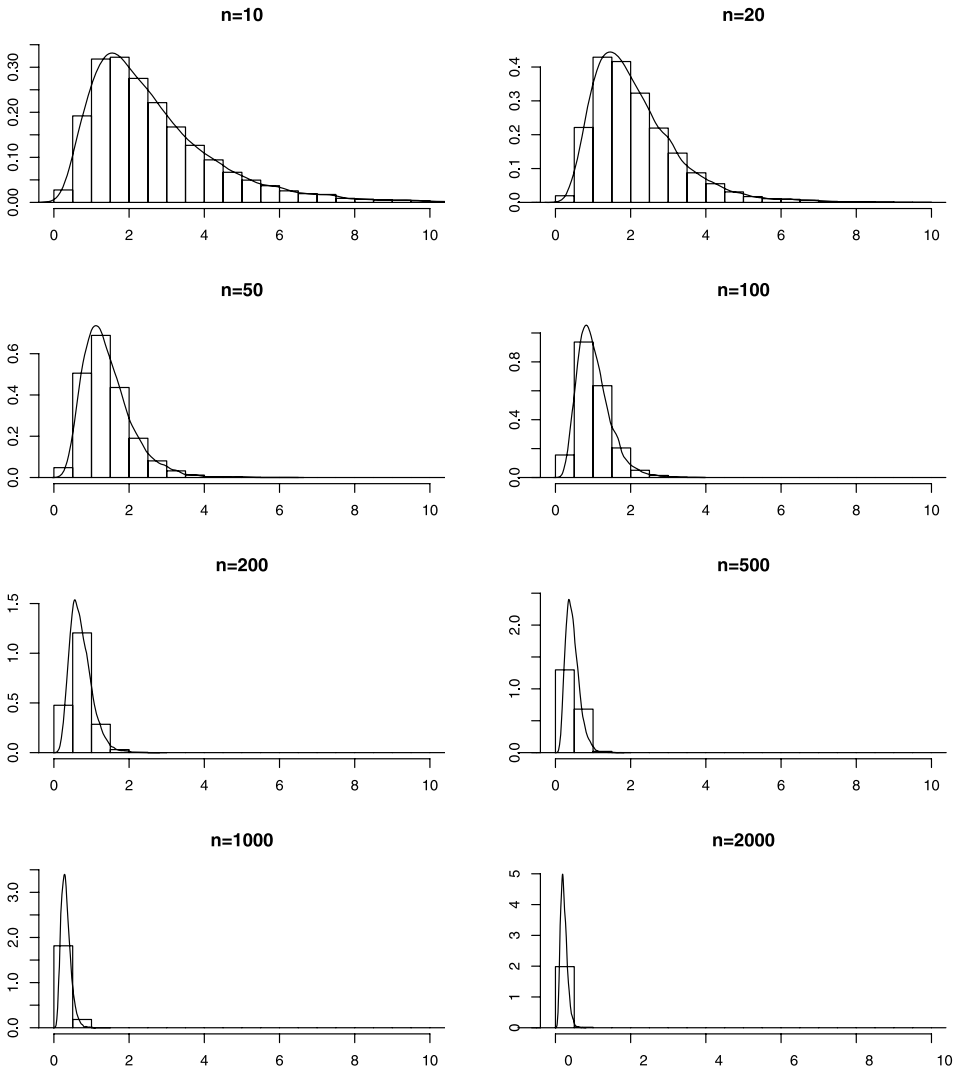

FIG. 5. *Standardized histograms overlaid by kernel density estimators of* $\sqrt{n}\widehat{\theta}_2$.

$\tau_2 = 09{:}40, \ldots, \tau_m = 16{:}00$ with $m = 78$. Denote by $X_i(t_{ij})$ the stock price on the ith day at the time t_{ij}, $j = 1, \ldots, n_i$ and $i = 1, \ldots, n$. The previous tick times on the ith day are defined as

$$\tau_{il} = \max\{t_{ij} : t_{ij} \le \tau_l, j = 1, \ldots, n_i\}, \qquad l = 1, \ldots, m.$$

The lth return on the ith day is then defined as $Z_{il} = \log\{X_i(\tau_{il})/X_i(\tau_{i,l-1})\}$.

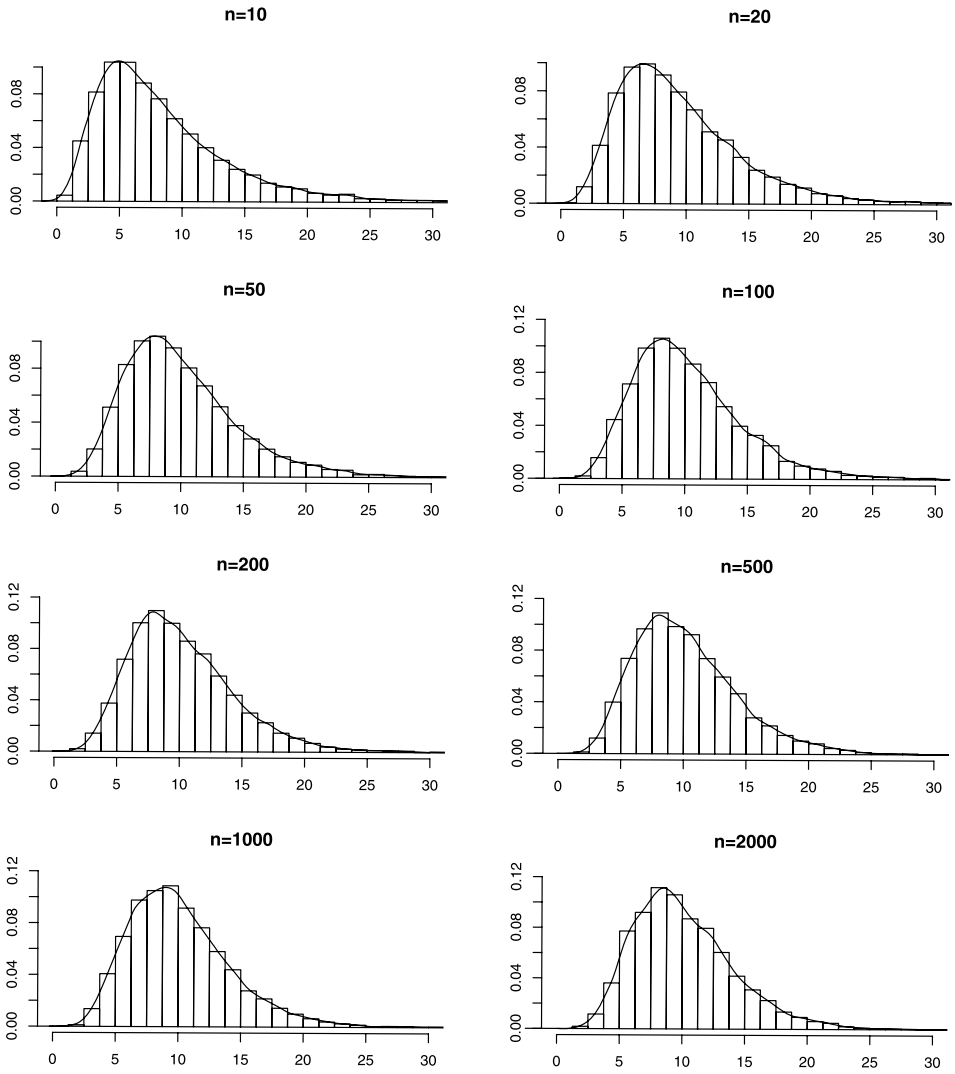

FIG. 6. *Standardized histograms overlaid by kernel density estimators of* $n\widehat{\theta}_2$.

We then estimate the intraday return densities using the standard kernel method

$$Y_i(u) = (mh_i)^{-1} \sum_{j=1}^{m} K\left(\frac{Z_{ij} - u}{h_i}\right), \qquad i = 1, \ldots, n, \tag{4.2}$$

where $K(u) = (\sqrt{2\pi})^{-1} \exp(-u^2/2)$ is a Gaussian kernel and h_i is a bandwidth. We set $\mathcal{I} = [-0.002, 0.002]$ as the support for $Y_i(\cdot)$. Let $\widehat{\sigma}_i$ be the sample standard deviation of $\{Z_{ij}, j = 1, \ldots, m\}$ and $\widehat{h}_i = 1.06\widehat{\sigma}_i m^{-1/5}$ be

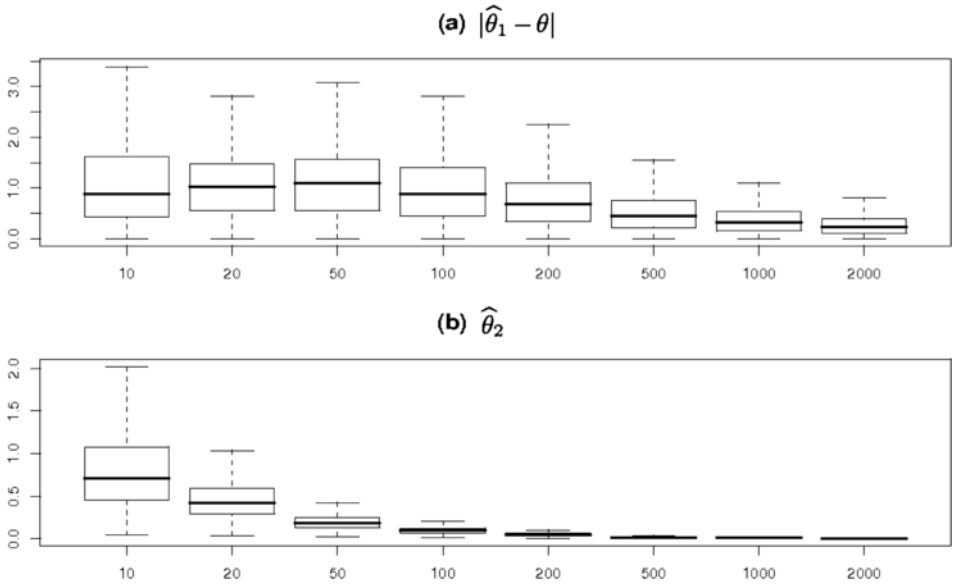

FIG. 7. *Boxplots of estimation errors:* (a) *Errors for nonzero eigenvalue* $|\widehat{\theta}_1 - \theta|$*;* (b) *Errors for zero-eigenvalue* $\widehat{\theta}_2$*. To add clarity to the display, the outliers are not plotted.*

Silverman's rule of thumb bandwidth choice for day i. Then for each i, we employ three levels of smoothness by setting h_i in (4.2) equal to $0.5\widehat{h}_i$, $\widehat{h}_i$ and $2\widehat{h}_i$. Figure 8 displays the observed densities for the first 8 days of the sample.

To identify the finite dimensionality of $Y_t(\cdot)$, we apply the methodology developed in this paper. We set $p = 5$ in (2.8). Figure 9 displays the estimated eigenvalues. With all three bandwidths used, the first two eigenvalues are much larger than the remaining ones. Furthermore, there is no clear cut-off from the third eigenvalue onwards. This suggests to take $\widehat{d} = 2$. The bootstrap tests, reported in Table 1, lend further support to this assertion. Indeed for all levels of smoothness adopted, the bootstrap test rejects the null $H_0 : \theta_2 = 0$ but cannot reject the hypothesis $\theta_j = 0$ for $j = 3, 4$ or 5. Note that it is implied by $\theta_3 = 0$ that $\theta_{3+k} = 0$ for $k \geq 1$. Indeed, we tested $\theta_{3+k} = 0$ only for illustrative purposes.

Table 2 contains the P-values from testing the hypothesis that the estimated loadings, $\widehat{\eta}_{tj}$ in (2.11) are white noise using the Ljung–Box–Pierce portmanteau test. Although we should interpret the results of this test with caution (see Remark 3 in Section 2.2.3), they provide further evidence that there is a considerable amount of dynamic structure in the two-dimensional subspace corresponding to the first two eigenvalues θ_1 and θ_2, and there is little or none dynamic structure in the directions corresponding to θ_3 and θ_4. Collating all the relevant findings, we comfortably set $\widehat{d} = 2$ in our analysis.

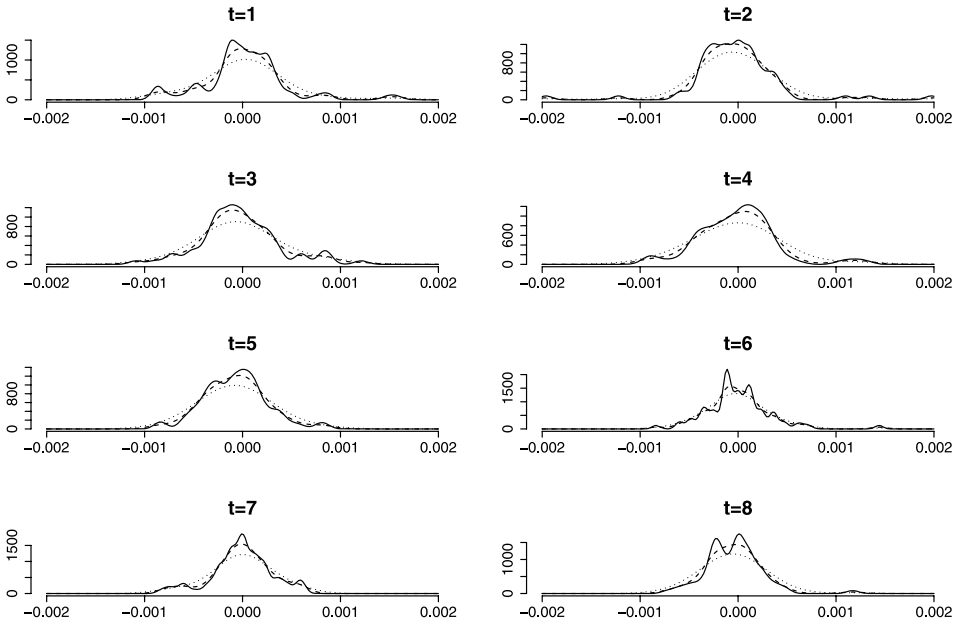

FIG. 8. *Estimated densities, $Y_i(\cdot)$, using bandwidths $h_i = \widehat{h}_i$ (solid lines), $0.5\widehat{h}_i$ (dashed lines) and $2\widehat{h}_i$ (dotted lines).*

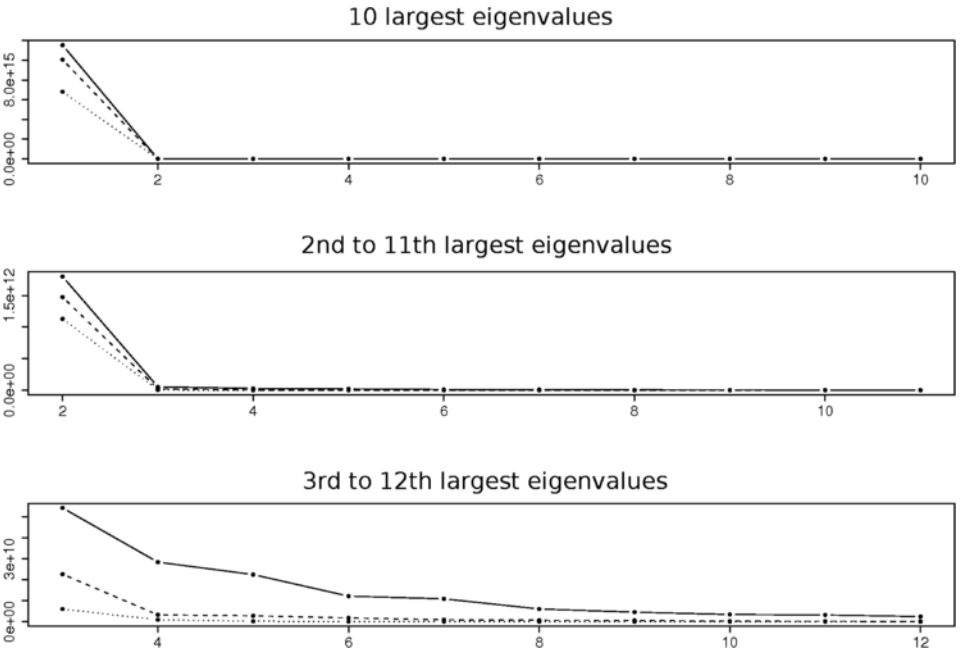

FIG. 9. *Estimated eigenvalues $\widehat{\theta}_j$ using bandwidths $h_t = 0.5\widehat{h}_t$ (solid lines), $\widehat{h}_t$ (dashed lines) and $2\widehat{h}_t$ (dotted lines).*

TABLE 1
P-values from applying the bootstrap test in Section 2.2.3 to the intraday return density example

	$h_t = 0.5\widehat{h}_t$	$h_t = \widehat{h}_t$	$h = 2\widehat{h}_t$
$H_0 : \theta_1 = 0$	0.00	0.00	0.00
$H_0 : \theta_2 = 0$	0.00	0.00	0.00
$H_0 : \theta_3 = 0$	0.35	0.15	0.18
$H_0 : \theta_4 = 0$	0.62	0.73	0.74
$H_0 : \theta_5 = 0$	0.68	0.91	0.93

Figure 10 displays the first $\widehat{d}(= 2)$ estimated eigenfunctions $\widehat{\psi}_j$ in (2.13). Although the estimated curves $Y_t(\cdot)$ in Figure 8 are somehow different for different bandwidths, the shape of the estimated eigenfunctions is insensitive to the choice of bandwidth.

Figure 11 displays time series plots of the estimated loadings $\widehat{\eta}_{t1}$ and $\widehat{\eta}_{tj}$. Again the estimated loadings with three levels of bandwidth are almost indistinguishable from each other. Furthermore, the ACF and PACF of the series $\widehat{\boldsymbol{\eta}}_{tj} = (\widehat{\eta}_{t1}, \widehat{\eta}_{t2})'$ are also virtually identical for all three choices of h. These graphics are displayed in Figures 12 and 13.

We now fit a VAR model to the estimated loadings, $\widehat{\boldsymbol{\eta}}_t$:

$$\widehat{\boldsymbol{\eta}}_t = \sum_{k=1}^{\tau} \mathbf{A}_k \widehat{\boldsymbol{\eta}}_{t-k} + \mathbf{e}_t. \tag{4.3}$$

Since the estimated loadings $\widehat{\boldsymbol{\eta}}_{tj}$, as defined in (2.11), have mean zero by construction, there is no intercept term in the model. We choose the order τ in (4.3) by minimizing the AIC. The AIC values for the order $\tau = 0, 1, \ldots, 10$ are given in Table 3. With all three bandwidths used, the AIC chooses $\tau = 3$, and the multivariate portmanteau test (with lag values 1, 3 and 5) of Li and

TABLE 2
P-values from testing the hypothesis $H_0 : \widehat{\eta}_{tj}$ is white noise using the Ljung–Box–Pierce portmanteau test. The test statistic is given by $Q_j = n(n+2)\sum_{k=1}^{q} s_j(k)^2/(n-k)$, where $s_j(k)$ is the sample autocorrelation of $\widehat{\eta}_{tj}$ at lag k. Under H_0, Q_j has an asymptotic χ^2_q-distribution

h_t	$0.5\widehat{h}_t$			$\widehat{h}_t$			$2\widehat{h}_t$		
q	1	3	5	1	3	5	1	3	5
$\widehat{\eta}_{t1}$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$\widehat{\eta}_{t2}$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$\widehat{\eta}_{t3}$	0.11	0.08	0.04	0.09	0.08	0.02	0.07	0.06	0.02
$\widehat{\eta}_{t4}$	0.05	0.25	0.28	0.22	0.33	0.47	0.53	0.56	0.63
$\widehat{\eta}_{t5}$	0.22	0.19	0.39	0.30	0.47	0.58	0.73	0.77	0.81

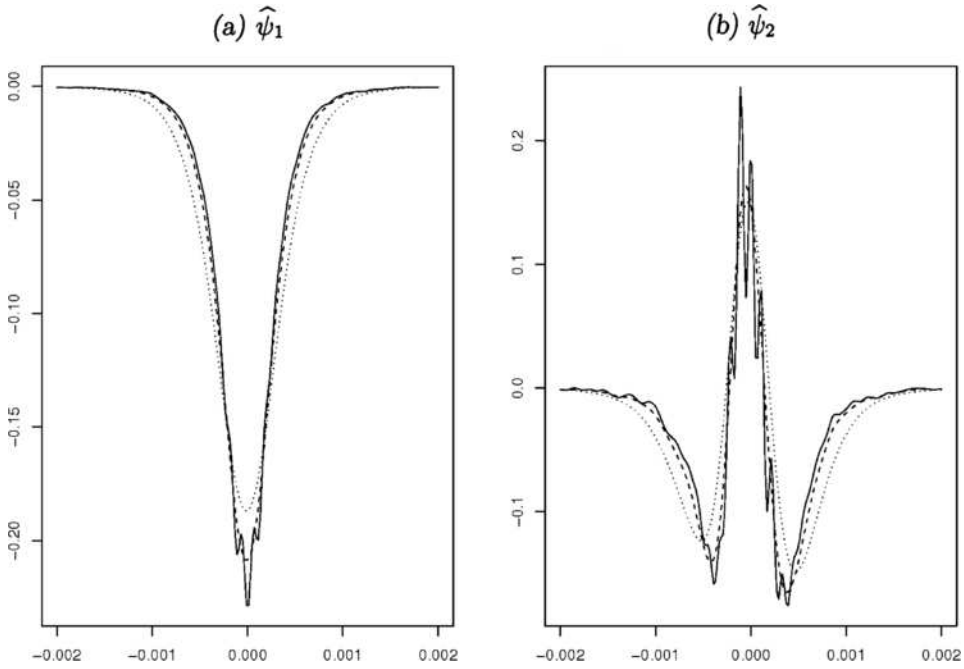

FIG. 10. *Estimated eigenfunctions* (a) $\widehat{\psi}_1$ *and* (b) $\widehat{\psi}_2$ *using bandwidths* $h_t = 0.5\widehat{h}_t$ *(solid lines),* $\widehat{h}_t$ *(dashed lines) and* $2\widehat{h}_t$ *(dotted lines).*

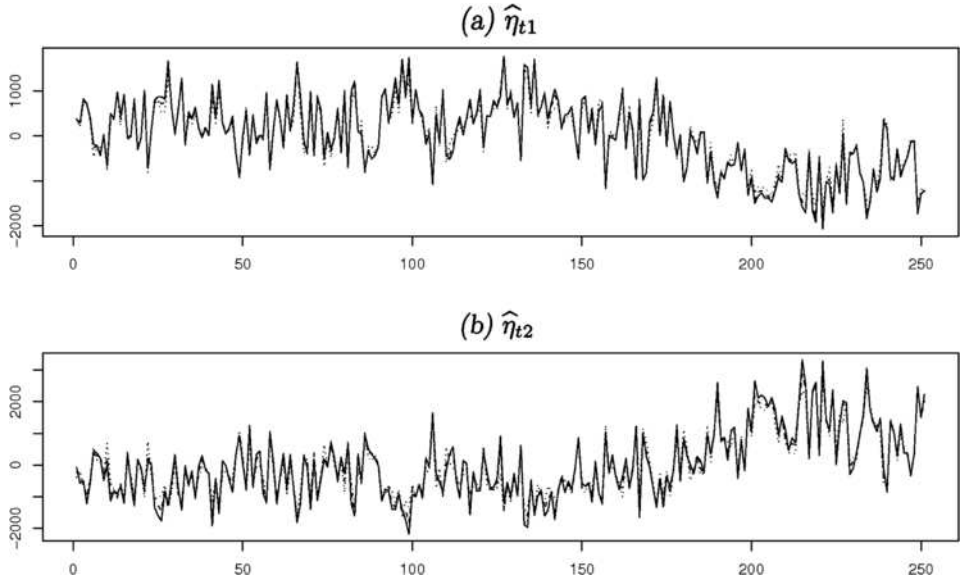

FIG. 11. *Estimated loadings* (a) $\widehat{\eta}_{t1}$ *and* (b) $\widehat{\eta}_{t2}$ *using bandwidths* $h_t = 0.5\widehat{h}_t$ *(solid lines),* $\widehat{h}_t$ *(dashed lines) and* $2\widehat{h}_t$ *(dotted lines).*

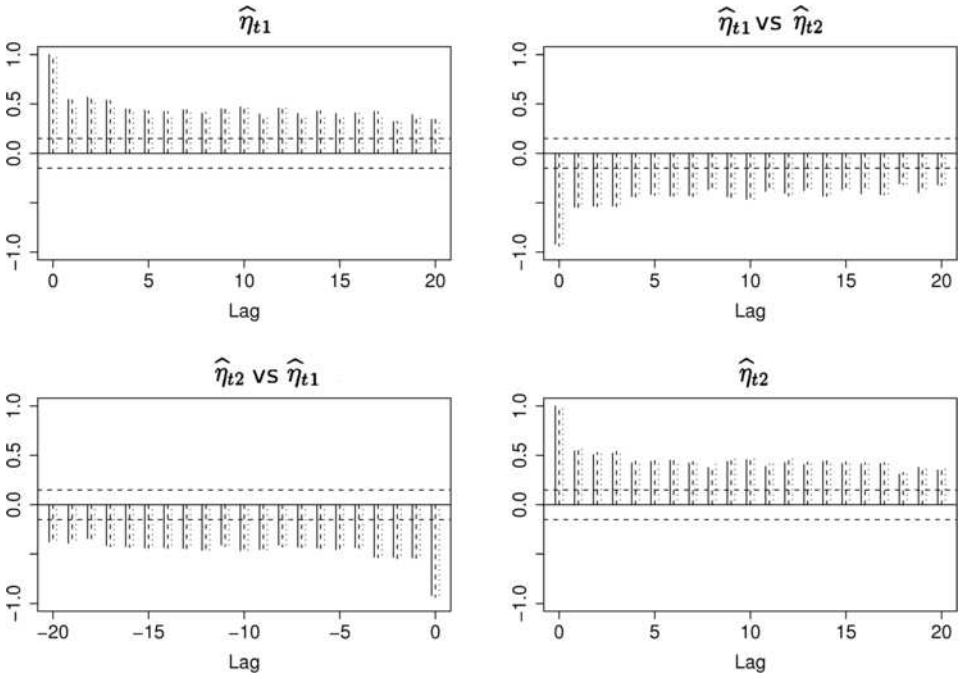

FIG. 12. *ACF of* $\widehat{\boldsymbol{\eta}}_{tj}$ *using bandwidths* $h_t = 0.5\widehat{h}_t$ *(solid lines),* $\widehat{h}_t$ *(dashed lines) and* $2\widehat{h}_t$ *(dotted lines).*

McLeod (1981) for the residual of the fitted VAR models are insignificant at the 10% level. The Yule–Walker estimates of the parameter matrices, $\mathbf{A}_k = (a_{k,ij})$ in (4.3), with the order $\tau = 3$ are given in Table 4.

To summarize, we found that the dynamic behavior of the IBM intraday return densities in 2006 was driven by two factors. These factors series are modeled well by a VAR(3) process. We note that with all the three levels of smoothness adopted in the initial density estimation, these conclusions were unchanged.

Finally, we make a cautionary remark on the implied true curves $X_t(\cdot)$ in the above analysis. We take the unknown true daily densities as $X_t(\cdot)$. We see those densities as random curves, as the distribution of the intraday returns tomorrow depends on the distributions of today, yesterday and so on, but is not entirely determined by them. Now in model (1.1), $E\{\varepsilon_t(u)\} = E\{Y_t(u)\} - E\{X_t(u)\} \not\equiv 0$. But this does not affect the analysis performed in identifying the dimensionality of the curves; see also Pan and Yao (2008). Note that (2.10) provides an alternative estimator for the true density $X_t(\cdot)$ based on the dynamic structure of the curve series. It can be used, for example, to forecast the density for tomorrow. However, an obvious normalization should be applied since we did not make use the constraint $\int_{\mathcal{I}} X_t(u)\,du = 1$ in constructing (2.10).

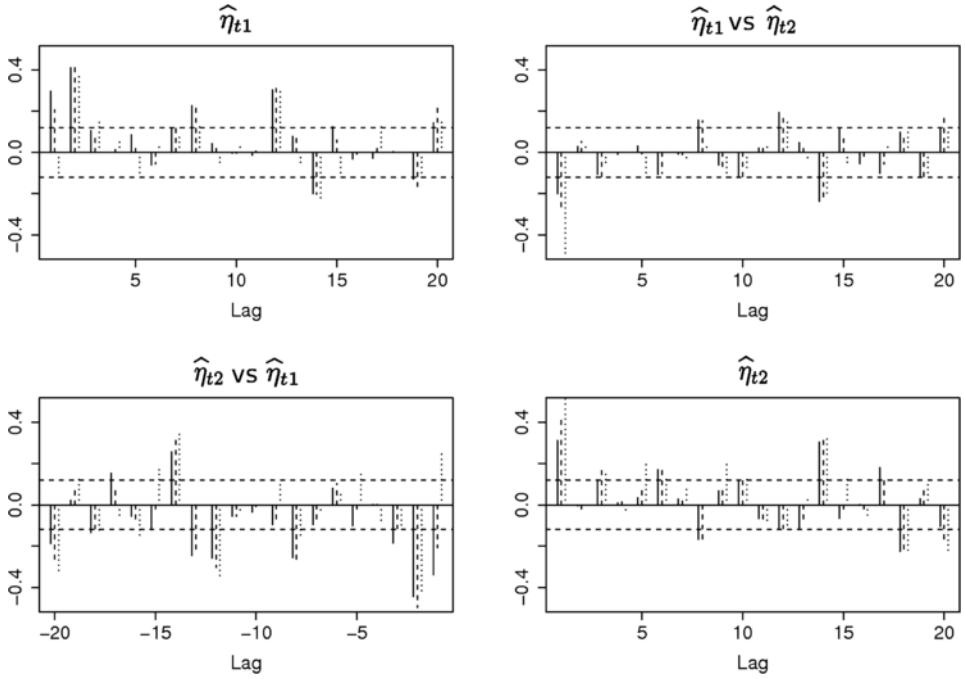

FIG. 13. *PACF of* $\widehat{\boldsymbol{\eta}}_{tj}$ *using bandwidths* $h_t = 0.5\widehat{h}_t$ *(solid lines),* $\widehat{h}_t$ *(dashed lines) and* $2\widehat{h}_t$ *(dotted lines).*

APPENDIX A

In this section, we provide the relevant background on operator theory used in this work. More detailed accounts may be found in Dunford and Schwartz (1988).

Let $\mathcal{H}$ be a real separable Hilbert space with respect to some inner product $\langle\cdot,\cdot\rangle$. For any $\mathcal{V} \subset \mathcal{H}$, the orthogonal complement of $\mathcal{V}$ is given by

$$\mathcal{V}^{\perp} = \{x \in \mathcal{H} : \langle x, y\rangle = 0, \forall y \in \mathcal{V}\}.$$

Note that $\mathcal{V}^{\perp\perp} = \overline{\mathcal{V}}$ where $\overline{\mathcal{V}}$ denotes the closure of $\mathcal{V}$. Clearly, if $\mathcal{V}$ is finite dimensional then $\mathcal{V}^{\perp\perp} = \mathcal{V}$.

Let L be a linear operator from $\mathcal{H}$ to $\mathcal{H}$. For $x \in \mathcal{H}$, denote by Lx the image of x under L. The adjoint of L is denoted by L^* and satisfies

$$\langle Lx, y\rangle = \langle x, L^* y\rangle, \qquad x, y \in \mathcal{H}. \tag{A.1}$$

L is said to be self adjoint if $L^* = L$ and nonnegative definite if

$$\langle Lx, x\rangle \geq 0 \qquad \forall x \in \mathcal{H}.$$

The image and null space of L are defined as $\mathrm{Im}(L) = \{y \in \mathcal{H} : y = Lx, x \in \mathcal{H}\}$ and $\mathrm{Ker}(L) = \{x \in \mathcal{H} : Lx = 0\}$, respectively. Note that $\mathrm{Ker}(L^*) = (\mathrm{Im}(L))^{\perp}$, $\mathrm{Ker}(L) = (\mathrm{Im}(L^*))^{\perp}$ and $\mathrm{Ker}(L^*) = \mathrm{Ker}(LL^*)$. We define the rank of L to be $r(L) = \dim(\mathrm{Im}(L))$ and we say that L is finite dimensional if $r(L) < \infty$.

TABLE 3
AIC values from fitting the VAR model in (4.3). The figures in this table have been centered at the minimum AIC value

	$\tau = 0$	$\tau = 1$	$\tau = 2$	$\tau = 3$	$\tau = 4$	$\tau = 5$
$h_t = 0.5\widehat{h}_t$	131.33	40.39	9.98	0.00	7.86	10.38
$h_t = \widehat{h}_t$	133.04	41.32	9.53	0.00	7.47	10.08
$h_t = 2\widehat{h}_t$	135.47	40.83	9.58	0.00	7.00	8.94

A linear operator L is said to be bounded if there exists some finite constant $\Delta > 0$ such that for all $x \in \mathcal{H}$

$$\|Lx\| < \Delta\|x\|,$$

where $\|\cdot\|$ is the norm induced on $\mathcal{H}$ by $\langle\cdot,\cdot\rangle$. We denote the space of bounded linear operators from $\mathcal{H}$ to $\mathcal{H}$ by $\mathcal{B} = \mathcal{B}(\mathcal{H}, \mathcal{H})$ and the uniform topology on $\mathcal{B}$ is defined by

$$\|L\|_{\mathcal{B}} = \sup_{\|x\| \le 1} \|Lx\|, \qquad L \in \mathcal{B}.$$

Note that all bounded linear operators are continuous, and the converse also holds.

An operator $L \in \mathcal{B}$ is said to be compact if there exists two orthonormal sequences $\{e_j\}$ in $\{f_j\}$ of $\mathcal{H}$ and a sequence of scalars $\{\lambda_j\}$ decreasing to zero such that

$$Lx = \sum_{j=1}^{\infty} \lambda_j \langle e_j, x\rangle f_j, \qquad x \in \mathcal{H},$$

or more compactly

$$L = \sum_{j=1}^{\infty} \lambda_j e_j \otimes f_j. \tag{A.2}$$

TABLE 4
Estimated parameter matrices $\mathbf{A}_k = (a_{k,ij})$ from fitting the VAR model in (4.3)

j	1			2		
h_t	$0.5\widehat{h}_t$	$\widehat{h}_t$	$2\widehat{h}_t$	$0.5\widehat{h}_t$	$\widehat{h}_t$	$2\widehat{h}_t$
$a_{1,1j}$	0.08	0.07	0.01	−0.14	−0.16	−0.22
$a_{1,2j}$	−0.08	−0.05	0.03	0.24	0.26	0.33
$a_{2,1j}$	0.35	0.39	0.38	0.06	0.09	0.08
$a_{2,2j}$	−0.36	−0.43	−0.43	−0.05	−0.10	−0.11
$a_{3,1j}$	0.08	0.05	0.02	−0.13	−0.15	−0.18
$a_{3,2j}$	−0.16	−0.13	−0.11	0.14	0.15	0.17

Note that if $\mathcal{H} = \mathcal{L}_2(\mathcal{I})$ equipped with the inner product defined in (2.1), then

$$(Lx)(u) = \sum_{j=1}^{\infty} \lambda_j \langle e_j, x\rangle f_j(u).$$

Clearly, $\mathrm{Im}(L) = \mathrm{sp}\{f_j : j \geq 1\}$ and $\mathrm{Ker}(L) = \mathrm{sp}\{e_j : j \geq 1\}^{\perp}$.

The Hilbert–Schmidt norm of a compact linear operator L is defined as $\|L\|_{\mathcal{S}} = (\sum_{j=1}^{\infty} \lambda_j^2)^{1/2}$. We will let $\mathcal{S}$ denote the space consisting of all the operators with a finite Hilbert–Schmidt or nuclear norm. Clearly, we have the inequalities $\|\cdot\|_{\mathcal{S}} \geq \|\cdot\|_{\mathcal{B}}$, and thus the inclusions $\mathcal{S} \subset \mathcal{B}$. Note that $\mathcal{B}$ is a Banach space when equipped with their respective norms. Furthermore, $\mathcal{S}$ is a Hilbert space with respect to the inner product

$$\langle L_1, L_2\rangle_{\mathcal{S}} = \sum_{i,j=1}^{\infty} \langle L_1 g_i, h_j\rangle\langle L_2 g_i, h_j\rangle, \qquad L_1, L_2 \in \mathcal{S},$$

where $\{g_i\}$ and $\{h_j\}$ are any orthonormal bases of $\mathcal{H}$.

APPENDIX B

In this section, we provide the proofs for the propositions in Section 2 and the theorems in Section 3. Throughout the proofs, we may use C to denote some (generic) positive and finite constant which may vary from line to line. We introduce some technical lemmas first.

LEMMA 1. *Let L be a finite-dimensional operator such that for some sequences of orthonormal vectors $\{e_j\}$, $\{f_j\}$, $\{g_j\}$ and $\{h_j\}$ and some sequences of decreasing scalars $\{\theta_j\}$ and $\{\lambda_j\}$, L admits the spectral decompositions $L = \sum_{j=1}^{d} \theta_j e_j \otimes f_j = \sum_{j=1}^{d'} \lambda_j g_j \otimes h_j$. Then it holds that $d' = d$.*

PROOF. Note that if $d \neq d'$ then both $\mathrm{Im}(L)$ and $\mathrm{Im}(L_k^*)$ will be of different dimensions under the alternative characterizations due to linear independence of $\{e_j\}$, $\{f_j\}$, $\{g_j\}$ and $\{h_j\}$. Thus, it must hold that $d = d'$. □

LEMMA 2. *Let L be a linear operator from $\mathcal{H}$ to $\mathcal{H}$, where $\mathcal{H}$ is a separable Hilbert space. Then it holds that $\overline{\mathrm{Im}(LL^*)} = \overline{\mathrm{Im}(L)}$.*

PROOF. Using the facts about inner product spaces and linear operators stated in Appendix A, we have

$$\begin{aligned}\overline{\mathrm{Im}(LL^*)} &= (\mathrm{Im}(LL^*))^{\perp\perp} = (\mathrm{Im}((LL^*)^*))^{\perp\perp} \\ &= (\mathrm{Ker}(LL^*))^{\perp} = (\mathrm{Ker}(L^*))^{\perp} \\ &= (\mathrm{Im}(L))^{\perp\perp} = \overline{\mathrm{Im}(L)},\end{aligned}$$

which concludes the proof. □

For the sake of the simplicity in presentation of the proofs, we adopt the the standard notation for Hilbert spaces. For any $f \in \mathcal{L}_2(\mathcal{I})$, we write $\|f\| = \sqrt{\langle f, f\rangle}$ [see (2.1)], and denote $M_k f \in \mathcal{L}_2(\mathcal{I})$ the image of f under the operator M_k in the sense that

$$(M_k f)(u) = \int_{\mathcal{I}} M_k(u, v) f(v)\, dv.$$

The operators $N_k, K, \widehat{M}_k$ and $\widehat{K}$ may be expressed in the same manner. Note now that the adjoint operator of M_k is

$$(M_k^* f)(u) = \int_{\mathcal{I}} M_k(v, u) f(v)\, dv.$$

See (A.1). Furthermore, $N_k = M_k M_k^*$ in the sense that $N_k f = M_k M_k^* f$; see (2.6). Similarly, $\widehat{K} = \sum_{k=1}^{p} \widehat{M}_k \widehat{M}_k^*$; see (2.9).

PROOF OF PROPOSITION 1. (i) To save notational burden, we set $k \equiv k_0$. We only need to show $\operatorname{Im}(N_k) = \mathcal{M}$. Since $N_k = M_k M_k^*$, it follows from Lemma 2 that $\operatorname{Im}(N_k) = \operatorname{Im}(M_k M_k^*) = \operatorname{Im}(M_k)$ as N_k and M_k are finite dimensional and thus their images are closed.

Now, recall from Section 2.1 that M_k may be decomposed as

$$M_k = \sum_{i,j=1}^{d} \sigma_{ij}^{(k)} \varphi_i \otimes \varphi_j. \tag{B.1}$$

See also (A.2). Thus, from (B.1), we may write

$$M_k = \sum_{i=1}^{d} \lambda_i^{(k)} \varphi_i \otimes \rho_i^{(k)}, \tag{B.2}$$

where

$$\rho_{ik} = \frac{\sum_{j=1}^{d} \sigma_{ij}^{(k)} \varphi_j}{\| \sum_{j=1}^{d} \sigma_{ij}^{(k)} \varphi_j \|}, \qquad \lambda_i^{(k)} = \left\| \sum_{j=1}^{d} \sigma_{ij}^{(k)} \varphi_i \right\|.$$

From (B.1), it is clear that $\operatorname{Im}(M_k) \subseteq \mathcal{M}$, which is finite dimensional. Thus, M_k is compact and therefore admits a spectral decomposition of the form

$$M_k = \sum_{j=1}^{d_k} \theta_j^{(k)} \psi_j^{(k)} \otimes \phi_j^{(k)} \tag{B.3}$$

with $(\phi_j^{(k)}, \psi_j^{(k)})$ forming the adjoint pair of singular functions of M_k corresponding to the singular value $\theta_j^{(k)}$. Clearly, $d_k \le d$. Thus, if $d_k < d$, $\operatorname{Im}(M_k) \subset \mathcal{M}$ since from (B.3), $\operatorname{Im}(M_k) = \operatorname{span}\{\phi_j^{(k)} : j = 1, \ldots, d_k\}$ and any subset of $d_k < d$ linearly independent elements in a d-dimensional space can only span a proper subset of the original space.

Now to complete the proof, we only need to show that the set of $\{\rho_j^{(k)}\}$ in (B.2) is linearly independent for some k. If this can be done, then we are in a position to apply Lemma 1. Let $\boldsymbol{\beta}$ be an arbitrary vector in $\mathbb{R}^d$ and put $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_d)'$ and $\boldsymbol{\rho}_k = (\rho_1^{(k)}, \ldots, \rho_d^{(k)})'$, then the linear independence of the set $\{\rho_i^{(k)}\}$ can easily be seen as the equation

$$\boldsymbol{\beta}\boldsymbol{\rho}_k = \boldsymbol{\beta}\boldsymbol{\Sigma}_k\boldsymbol{\varphi} = 0$$

has a nontrivial solution if and only if $\boldsymbol{\beta}\boldsymbol{\Sigma}_k = \mathbf{0}$. However, since $\boldsymbol{\Sigma}_k$ is of full rank by assumption, it follows that it is invertible and the only solution is the trivial one $\boldsymbol{\beta} = \mathbf{0}$. Thus, Lemma 1 implies $d_k = d$ and the result follows from noting that any linearly independent set of d elements in a d-dimensional vector space forms a basis for that space.

(ii) Similar to the proof of part (i) above, we only need to show $\operatorname{Im}(K) = \mathcal{M}$. Note that for any $f \in L_2(\mathcal{I})$, $\langle M_k M_k^* f, f\rangle = \langle M_k^* f, M_k^* f\rangle = \|M_k^* f\|^2 \geq 0$, thus the composition $N_k = M_k M_k^*$ is nonnegative definite which implies that K is also nonnegative definite. Therefore, $\operatorname{Im}(K) = \bigcup_{k=1}^p \operatorname{Im}(N_k)$. From here, the result given in part (i) of the proposition concludes the proof. □

Proof of Proposition 2. Let $\widehat{\theta}_j$ be a nonzero eigenvalue of $\mathbf{K}^*$, and $\boldsymbol{\gamma}_j = (\gamma_{1j}, \ldots, \gamma_{n-p,j})'$ be the corresponding eigenvector, that is, $\mathbf{K}^*\boldsymbol{\gamma}_j = \boldsymbol{\gamma}_j\widehat{\theta}_j$. Writing this equation component by component, we obtain that

$$\frac{1}{(n-p)^2}\sum_{i,s=1}^{n-p}\sum_{k=1}^{p}\langle Y_{t+k} - \bar{Y}, Y_{s+k} - \bar{Y}\rangle\langle Y_s - \bar{Y}, Y_i - \bar{Y}\rangle\gamma_{ij} = \gamma_{tj}\widehat{\theta}_j \tag{B.4}$$

for $t = 1, \ldots, n-p$; see (2.12). For $\widehat{\psi}_j$ defined in (2.13),

$$\begin{aligned}
(\widehat{K}\widehat{\psi}_j)(u) &= \int_{\mathcal{I}} \widehat{K}(u,v)\widehat{\psi}_j(v)\,dv \\
&= \frac{1}{(n-p)^2}\sum_{t,s=1}^{n-p}\sum_{k=1}^{p}\{Y_t(u) - \bar{Y}(u)\}\langle Y_s - \bar{Y}, \widehat{\psi}_j\rangle \\
&\qquad\qquad \times \langle Y_{t+k} - \bar{Y}, Y_{s+k} - \bar{Y}\rangle \\
&= \frac{1}{(n-p)^2}\sum_{t,s,i=1}^{n-p}\sum_{k=1}^{p}\{Y_t(u) - \bar{Y}(u)\}\gamma_{ij}\langle Y_s - \bar{Y}, Y_i - \bar{Y}\rangle \\
&\qquad\qquad \times \langle Y_{t+k} - \bar{Y}, Y_{s+k} - \bar{Y}\rangle;
\end{aligned}$$

see (2.9). Plugging (B.4) into the right-hand side of the above expression, we obtain that

$$(\widehat{K}\widehat{\psi}_j)(u) = \sum_{t=1}^{n-p}\{Y_t(u) - \bar{Y}(u)\}\gamma_{tj}\widehat{\theta}_j = \widehat{\psi}_j(u)\widehat{\theta}_j,$$

that is, $\widehat{\psi}_j$ is an eigenfunction of $\widehat{K}$ corresponding to the eigenvalue $\widehat{\theta}_j$. □

As we shall see, the operator $\widehat{K} = \sum_{k=1}^{p} \widehat{M}_k \widehat{M}_k^*$ may be written as a functional of empirical distributions of Hilbertian random variables. Thus, we require an auxiliary result to deal with this form of process. To this end, we extend the V-statistic results of Sen (1972) to the setting of Hilbertian valued random variables. Further details about V-statistics may be found in Lee (1990).

Let $\mathcal{H}$ be a real separable Hilbert space with norm $\|\cdot\|$ generated by an inner product $\langle\cdot,\cdot\rangle$. Let $X_t \in \mathcal{X}$ be a sequence of strictly stationary and Hilbertian random variables whose distribution functions will be denoted by $P(x), x \in \mathcal{H}$. Note that the spaces $\mathcal{X}$ and $\mathcal{H}$ may differ. Let $\phi : \mathcal{X}^m \to \mathcal{H}$ be Bochner integrable and symmetric in each of its $m(\geq 2)$ arguments. Now consider the functional

$$\theta(P) = \int_{\mathcal{X}^m} \phi(x_1, \ldots, x_m) \prod_{j=1}^{m} P(dx_j),$$

defined over $\mathcal{P} = \{P : \|\theta(P)\| < \infty\}$. As an estimator of $\theta(P)$, consider the V-statistic defined by

$$V_n = n^{-m} \sum_{i_1=1}^{n} \cdots \sum_{i_m=1}^{n} \phi(X_{i_1}, \ldots, X_{i_m}).$$

Now for $c = 0, 1, \ldots, m$, we define the functions

$$\phi_c(x_1, \ldots, x_c) = \int_{\mathcal{X}^{m-c}} \phi(x_1, \ldots, x_c, x_{c+1}, \ldots, x_m) \prod_{j=c+1}^{m} P(dx_j)$$

and

$$g_c(x_1, \ldots, x_c) = \sum_{d=0}^{c} (-1)^{c-d} \sum_{1 \leq j_1 < \cdots < j_d \leq c} \phi_d(X_{j_1}, \ldots, X_{j_d}).$$

In order to construct the canonical decomposition of V_n, we use Dirac's δ-measure to define the empirical measure P_n as follows:

$$P_n(A) = n^{-1}(\delta_{X_1}(A) + \cdots + \delta_{X_n}(A)), \qquad A \in \mathcal{X}.$$

Then for $c = 1, \ldots, m$, we set

$$V_{nc} = \int_{X^c} \phi_c(x_1, \ldots, x_c) \prod_{j=1}^{c} (P_n(dx_j) - P(dx_j))$$

$$= n^{-c} \sum_{i_1=1}^{n} \cdots \sum_{i_c=1}^{n} g_c(X_{i_1}, \ldots, X_{i_c}),$$

then we have

$$V_n - \theta(P) = \sum_{c=1}^{m} \binom{m}{c} V_{nc}. \tag{B.5}$$

In particular, note that

$$V_{n1} = \frac{1}{n}\sum_{i=1}^{n} g_1(X_i).$$

Decomposition (B.5) is the Hoeffding representation of the statistic V_n. It plays a central role in the proof of Lemma 3 below. We are now in a position to state some regularity conditions which form the basis of the result.

- A1. $\{X_t\}$ is strictly stationary and ψ-mixing with ψ-mixing coefficients satisfying the condition $\sum_{l=1}^{\infty} l^{m-1}\psi^{1/2}(l) < \infty$.
- A2. $\int_{\mathcal{X}^m} \|\phi(x_1,\ldots,x_m)\|^2 \prod_{j=1}^{m} P(dx_j) < \infty$.
- A3. $E\|g_1(X_1)\|^2 + 2\sum_{k=2}^{\infty} E\langle g_1(X_1), g_1 X_k\rangle \neq 0$.

LEMMA 3. *Let conditions* A1–A3 *hold. Then for* $c = 1,\ldots,m$ *it holds that* $E\|V_{nc}\|^2 = O(n^{-c})$.

PROOF. We make use of (B.5). Let $\{e_j : j \geq 1\}$ be an orthonormal basis of $\mathcal{H}$. Then

$$E\|V_{nc}\|^2 = \sum_{j=1}^{\infty} E\langle e_j, V_{nc}\rangle^2, \tag{B.6}$$

where $\langle e_j, V_{nc}\rangle$ is the $\mathbb{R}$ valued V-statistic

$$\langle e_j, V_{nc}\rangle = n^{-c}\sum_{i_1=1}^{n}\cdots\sum_{i_c=1}^{n}\langle e_j, g_c(X_{i_1},\ldots,X_{i_c})\rangle.$$

Now under conditions A1–A3, Lemma 3.3 in Sen (1972) yields

$$E\langle e_j, V_{nc}\rangle^2 \leq Cn^{-c}\int_{\mathcal{X}^c}\langle e_j, \phi_c(x_1,\ldots,x_c)\rangle^2 \prod_{j=1}^{c} P(dx_j) \tag{B.7}$$

for all $j \geq 1$. Now inserting the estimate in (B.7) into (B.6) yields

$$\begin{aligned}
E\|V_{nc}\|^2 &\leq Cn^{-c}\sum_{j=1}^{\infty}\int_{\mathcal{X}^c}\langle e_j, \phi_c(x_1,\ldots,x_c)\rangle^2 \prod_{j=1}^{c} P(dx_j)\\
&\leq Cn^{-c}\int_{\mathcal{X}^c}\|\phi_c(x_1,\ldots,x_c)\|^2 \prod_{j=1}^{c} P(dx_j)\\
&\leq Cn^{-c}\sum_{j=1}^{\infty}\int_{\mathcal{X}^c}\|\phi(x_1,\ldots,x_m)\|^2 \prod_{j=1}^{m} P(dx_j)\\
&= O(n^{-c})
\end{aligned}$$

as required. □

PROOF OF THEOREM 1. (i) Since p is fixed and finite, we may set $n \equiv n-p$. Let $Z_{tk} = (Y_t - \mu) \otimes (Y_{t+k} - \mu) \in \mathcal{S}$. Now consider the kernel $\rho : \mathcal{S} \times \mathcal{S} \to \mathcal{S}$ given by

$$\rho(A, B) = AB^*, \qquad A, B \in \mathcal{S}. \tag{B.8}$$

Now note that from (B.8),

$$\widehat{M}_k \widehat{M}_k = n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} \rho(Z_{ik}, Z_{jk}),$$

which in light of the preceding discussion is simply a $\mathcal{S}$ valued von Mises functional. Then $d \geq 1$ it holds that $M_k \neq 0$, an application of Lemma 3 yields

$$E\|\widehat{M}_k \widehat{M}_k^* - M_k M_k^*\|_{\mathcal{S}}^2 = O(n^{-1}). \tag{B.9}$$

Note that if $d = 0$, the rate in (B.9) would be n^{-2}, that is, the kernel ρ would possess the property of first order degeneracy. Now by (B.9) and the Chebyshev inequality, we have

$$\|\widehat{K} - K\|_{\mathcal{S}} \leq \sum_{k=1}^{p} \|\widehat{M}_k \widehat{M}_k - M_k M_k^*\|_{\mathcal{S}} = O_p(n^{-1/2}).$$

(ii) Given $\|\widehat{K} - K\|_{\mathcal{S}} = O_p(n^{-1/2})$, Lemma 4.2 in Bosq (2000) implies the $\sup_{j \geq 1} |\widehat{\theta}_j - \theta_j| \leq \|\widehat{K} - K\|_{\mathcal{S}} = O_p(n^{-1/2})$. Condition C3 ensures that ψ_j is an identifiable statistical parameter for $j = 1, \ldots, d$. From here, Lemma 4.3 in Bosq (2000) implies $\|\widehat{\psi}_j - \psi_j\| \leq C\|\widehat{K} - K\|_{\mathcal{S}} = O_p(n^{-1/2})$.

(iii) First, note that by Lemma 3 we have

$$E\|\widehat{M}_k \widehat{M}_k^* - \widehat{M}_k M_k^*\|_{\mathcal{S}}^2 = O(n^{-2}). \tag{B.10}$$

Put $\widetilde{K} = \sum_{k=1}^{p} \widehat{M}_k M_k$. Then by (B.10) and the Chebyshev inequality, we have

$$\|\widehat{K} - \widetilde{K}\|_{\mathcal{S}} \leq \sum_{k=1}^{p} \|\widehat{M}_k \widehat{M}_k^* - \widehat{M}_k M_k^*\|_{\mathcal{S}} = O_p(n^{-1}). \tag{B.11}$$

The estimate in (B.11) will prove to be crucial in deriving the results for $\widehat{\theta}_j$ when $j \geq d+1$.

Now, extend $\psi_1, \ldots, \psi_d$ to a complete orthonormal basis of $\mathcal{H}$. Then it holds that

$$\sum_{j=1}^{n} \widehat{\theta}_j = \sum_{j=1}^{\infty} \langle \psi_j, \widehat{K} \psi_j \rangle, \tag{B.12}$$

and by recalling that $\theta_j = 0$ for $j > d$

$$\sum_{j=1}^{d} \theta_j = \sum_{j=1}^{d} \langle \psi_j, K \psi_j \rangle. \tag{B.13}$$

Note that $\operatorname{span}\{\psi_j : j > d\} = \mathcal{M}^{\perp}$ and $K\psi_j = 0$ for all $j > d$ since $\operatorname{Ker}(K) = \mathcal{M}^{\perp}$. Thus, from (B.12) and (B.13), we have

$$\sum_{j=1}^{n} \widehat{\theta}_j - \theta_j = \sum_{j=1}^{\infty} \langle \psi_j, (\widehat{K} - K)\psi_j \rangle. \tag{B.14}$$

Now we will show that

$$\widehat{\theta}_j - \theta_j = \langle \psi_j, (\widehat{K} - K)\psi_j \rangle + O_p(n^{-1}), \qquad j = 1, \ldots, d. \tag{B.15}$$

Let $K_j = \langle \psi_j, (\widehat{K} - K)\widehat{\psi}_j \rangle$. Then using the relations $K\psi_j = \theta_j\psi_j$ and $\widehat{K}\widehat{\psi}_j = \widehat{\theta}_j\widehat{\psi}_j$ along with the fact that K is self adjoint, we have

$$\begin{aligned} |K_j - (\widehat{\theta}_j - \theta_j)| &= |\langle \psi_j, \widehat{K}\widehat{\psi}_j \rangle - \langle K\psi_j, \widehat{\psi}_j \rangle - (\widehat{\theta}_j - \theta_j)| \\ &= |(\widehat{\theta}_j - \theta_j)(\langle \psi_j, \widehat{\psi}_j \rangle - 1)| \\ &= |\widehat{\theta}_j - \theta_j||\langle \psi_j, \widehat{\psi}_j \rangle - 1|. \end{aligned} \tag{B.16}$$

Note that

$$|\langle \psi_j, \widehat{\psi}_j \rangle - 1| = |\langle \psi_j, \widehat{\psi}_j - \psi_j \rangle| \le \|\psi_j\|\|\widehat{\psi}_j - \psi_j\| = \|\widehat{\psi}_j - \psi_j\|. \tag{B.17}$$

Thus, from the results in (b) above (B.16) and (B.17), we have $|K_j - (\widehat{\theta}_j - \theta_j)| \le |\widehat{\theta}_j - \theta_j|\|\widehat{\psi}_j - \psi_j\| = O_p(n^{-1})$ for $j = 1, \ldots, d$.

Next, we have

$$\begin{aligned} |\langle \psi_j, (\widehat{K} - K)\psi_j \rangle - K_j| &= |\langle \psi_j - \widehat{\psi}_j, (\widehat{K} - K)\psi_j \rangle| \\ &\le \|\psi_j - \widehat{\psi}_j\|\|(\widehat{K} - K)\psi_j\| \\ &\le \|\psi_j - \widehat{\psi}_j\|\|\widehat{K} - K\|_{\mathcal{S}}, \end{aligned}$$

from which the results in (i) and (ii) $|\langle \psi_j, (\widehat{K} - K)\psi_j \rangle - K_j| = O_p(n^{-1})$, thus proving (B.15).

Now from (B.15) we have

$$\sum_{j=1}^{d} \widehat{\theta}_j - \theta_j = \sum_{j=1}^{d} \langle \psi_j, (\widehat{K} - K)\psi_j \rangle + O_p(n^{-1}),$$

and thus from (B.11) and (B.14)

$$\begin{aligned} \sum_{j=d+1}^{n} \widehat{\theta}_j &= \sum_{j=d+1}^{\infty} \langle \psi_j, (\widehat{K} - K)\psi_j \rangle + O_p(n^{-1}) \\ &= \sum_{j=d+1}^{\infty} \langle \psi_j, (\widetilde{K} - K)\psi_j \rangle + O_p(n^{-1}). \end{aligned}$$

By noting that $\psi_j \in \mathcal{M}^\perp$ for $j \geq d+1$ and $\mathrm{Ker}(M_k) = \mathrm{Ker}(\widetilde{K}) = \mathrm{Ker}(K) = \mathcal{M}^\perp$, it holds that $\sum_{j=d+1}^{\infty} \langle \psi_j, (\widetilde{K} - K)\psi_j \rangle = 0$. Thus, $\sum_{j=d+1}^{n} \widehat{\theta}_j = O_p(n^{-1})$ and the result follows from noting that $\widehat{\theta}_i \leq \sum_{j=d+1}^{n} \widehat{\theta}_j$ for $i = 1, \ldots, d$.

(iv) Let $\Pi_{\mathcal{M}}$ and $\Pi_{\mathcal{M}^\perp}$ denote the projection operators onto $\mathcal{M}$ and $\mathcal{M}^\perp$, respectively. Since $x = \Pi_{\mathcal{M}}(x) + \Pi_{\mathcal{M}^\perp}(x)$ for any $x \in \mathcal{L}_2(\mathcal{I})$, we have

$$\|\Pi_{\mathcal{M}}(\widehat{\psi}_i)\|^2 = \|\widehat{\psi}_i - \Pi_{\mathcal{M}^\perp}(\widehat{\psi}_i)\|^2 = \sum_{j=1}^{d} \langle \widehat{\psi}_i, \psi_j \rangle^2 \tag{B.18}$$

for all $i \geq 1$. Now note that for $i \geq d+1$

$$\begin{aligned} \|K(\widehat{\psi}_i)\| &= \|(K - \widehat{K})(\widehat{\psi}_i) + \widehat{\psi}_i \widehat{\theta}_i\| \\ &\leq \|(K - \widehat{K})(\widehat{\psi}_i)\| + |\widehat{\theta}_i| \|\widehat{\psi}_i\| \\ &\leq 2\|K - \widehat{K}\|_{\mathcal{B}}, \end{aligned} \tag{B.19}$$

where the final inequality follows from the definition of $\|\cdot\|_{\mathcal{B}}$ and Lemma 4.2 in Bosq (2000) by noting that $\theta_i = 0$ for all $i \geq d+1$.

Next, we have for $i \geq d+1$

$$\begin{aligned} \|K(\widehat{\psi}_i)\|^2 &= \sum_{j=1}^{\infty} \langle K(\widehat{\psi}_i), \psi_j \rangle^2 = \sum_{j=1}^{\infty} \theta_j^2 \langle \widehat{\psi}_i, \psi_j \rangle^2 \\ &= \sum_{j=1}^{d} \theta_j^2 \langle \widehat{\psi}_i, \psi_j \rangle^2 \geq \theta_d^2 \sum_{j=1}^{d} \langle \widehat{\psi}_i, \psi_j \rangle^2, \end{aligned} \tag{B.20}$$

since $\theta_1 > \cdots > \theta_d$. Combining (B.18), (B.19) and (B.20) yields

$$\|\Pi_{\mathcal{M}}(\widehat{\psi}_{d_0+1})\|^2 = \|\widehat{\psi}_{d_0+1} - \Pi_{\mathcal{M}^\perp}(\widehat{\psi}_{d_0+1})\|^2 \leq C\|K - \widehat{K}\|_{\mathcal{B}},$$

from which (i) yields the result. □

LEMMA 4. *The function D defined in (3.2) is a well-defined distance measure on $\mathcal{Z}_D$.*

PROOF. Nonnegativity, symmetry and the identity of indiscernibles are obvious. It only remains to prove the subadditivity property. For any $L \in \mathcal{S}$, note that $\|L\|_{\mathcal{S}} = \sqrt{\mathrm{tr}(L^*L)}$, where tr denotes the trace operator. Now, for any $\mathcal{X}_i \in \mathcal{Z}$, $i = 1, 2, 3$, let $\Pi_{\mathcal{X}_i}$ denote its corresponding d-dimensional projection operators defined as follows:

$$\Pi_{\mathcal{X}_i} = \sum_{j=1}^{d} \zeta_{ij} \otimes \zeta_{ij},$$

where $\{\zeta_{ij} : j = 1, \ldots, d\}$ is some orthonormal basis of $\mathcal{X}_i$. Now the triangle inequality for the Hilbert–Schmidt norm yields

$$\|\Pi_{\mathcal{X}_1} - \Pi_{\mathcal{X}_3}\|_{\mathcal{S}} \le \|\Pi_{\mathcal{X}_1} - \Pi_{\mathcal{X}_2}\|_{\mathcal{S}} + \|\Pi_{\mathcal{X}_2} - \Pi_{\mathcal{X}_3}\|_{\mathcal{S}}.$$

Since the projection operators are self adjoint, we have

$$\begin{aligned}
&\sqrt{\operatorname{tr}(\Pi_{\mathcal{X}_1}^2) + \operatorname{tr}(\Pi_{\mathcal{X}_3}^2) - 2\operatorname{tr}(\Pi_{\mathcal{X}_1}\Pi_{\mathcal{X}_3})} \\
&\quad \le \sqrt{\operatorname{tr}(\Pi_{\mathcal{X}_1}^2) + \operatorname{tr}(\Pi_{\mathcal{X}_2}^2) - 2\operatorname{tr}(\Pi_{\mathcal{X}_1}\Pi_{\mathcal{X}_2})} \\
&\qquad + \sqrt{\operatorname{tr}(\Pi_{\mathcal{X}_2}^2) + \operatorname{tr}(\Pi_{\mathcal{X}_3}^2) - 2\operatorname{tr}(\Pi_{\mathcal{X}_2}\Pi_{\mathcal{X}_3})}.
\end{aligned}$$

Now $\operatorname{tr}(\Pi_{\mathcal{X}_i}^2) = \operatorname{tr}(\Pi_{\mathcal{X}_i}) = d$ and $\operatorname{tr}(\Pi_{\mathcal{X}_i}\Pi_{\mathcal{X}_j}) = \sum_{k,l=1}^{d} \langle \zeta_{ik}, \zeta_{jl} \rangle^2$ for $i, j = 1, 2, 3$. These last facts along with the definition of D in (3.2) give

$$D(\mathcal{X}_1, \mathcal{X}_3) \le D(\mathcal{X}_1, \mathcal{X}_2) + D(\mathcal{X}_2, \mathcal{X}_3),$$

which concludes the proof. □

PROOF OF THEOREM 2. From the definition of D in (3.2), note that

$$\sqrt{2d} D(\widehat{\mathcal{M}}, \mathcal{M}) = \|\Pi_{\widehat{\mathcal{M}}} - \Pi_{\mathcal{M}}\|_{\mathcal{S}}, \tag{B.21}$$

where $\Pi_{\widehat{\mathcal{M}}} = \sum_{j=1}^{d} \widehat{\psi}_j \otimes \widehat{\psi}_j$ and $\Pi_{\mathcal{M}} = \sum_{j=1}^{d} \phi_j \otimes \phi_j$ with $\phi_1, \ldots, \phi_d$ forming any orthonormal basis of $\mathcal{M}$. Now if $\Pi_{\mathcal{M}}^1$ and $\Pi_{\mathcal{M}}^2$ are any projection operators onto $\mathcal{M}$, then by virtue of Lemma 4 it holds that $\|\Pi_{\mathcal{M}}^1 - \Pi_{\mathcal{M}}^2\|_{\mathcal{S}} = \sqrt{2d} D(\mathcal{M}, \mathcal{M}) = 0$. Thus, we may proceed as if $\Pi_{\mathcal{M}}$ in (B.21) was formed with eigenfunctions of K, that is, $\phi_j = \psi_j$ for $j = 1, \ldots, d$.

Now, we have

$$\left\| \sum_{j=1}^{d} \widehat{\psi}_j \otimes \widehat{\psi}_j - \sum_{j=1}^{d} \psi_j \otimes \psi_j \right\|_{\mathcal{S}} \le \sum_{j=1}^{d} \|\widehat{\psi}_j \otimes \widehat{\psi}_j - \psi_j \otimes \psi_j\|_{\mathcal{S}}, \tag{B.22}$$

that is, $\widehat{\psi}_j \otimes \widehat{\psi}_j$ (resp., $\psi_j \otimes \psi_j$) is the projection operator onto the eigensubspace generated by $\widehat{\theta}_j$ (resp., θ_j). Now by part (i) of Theorem 1, $\|\widehat{K} - K\|_{\mathcal{S}} = O_p(n^{-1/2})$. Thus, Theorem 2.2 in Mas and Menneteau (2003) implies that $\|\widehat{\psi}_j \otimes \widehat{\psi}_j - \psi_j \otimes \psi_j\|_{\mathcal{S}} = O_p(n^{-1/2})$ for $j = 1, \ldots, d$. This last fact along with (B.21) and (B.22) yield $D(\widehat{\mathcal{M}}, \mathcal{M}) = O_p(n^{-1/2})$. □

PROOF OF THEOREM 3. We first note that from (B.9), the triangle inequality and the c_r inequality, we have

$$E\|\widehat{K} - K\|_{\mathcal{S}}^2 = O(n^{-1}). \tag{B.23}$$

As $\widehat{\theta}_1 \geq \widehat{\theta}_2 \geq \cdots \geq 0$ (with strict inequality holding with probability one), it holds that $\{\widehat{d} > d\} = \{\widehat{\theta}_{d+1} > \epsilon\}$. Now since $\theta_{d+1} = 0$, it holds that $\widehat{\theta}_{d+1} = |\widehat{\theta}_{d+1} - \theta_{d+1}| \leq \|\widehat{K} - K\|_{\mathcal{S}}$ by Lemma 4.2 in Bosq (2000). Collecting these last few facts and applying the Chebyshev inequality yields

$$P(\widehat{d} > d) \leq \epsilon^{-2} E\|\widehat{K} - K\|_{\mathcal{S}}^2 = O((\epsilon^2 n)^{-1}) \tag{B.24}$$

by (B.23). Next, we turn to $P(\widehat{d} < d)$. Due to the ordering of the eigenvalues, it holds that $\{\widehat{d} < d\} = \{\widehat{\theta}_{d-1} < \epsilon\}$. Therefore,

$$\begin{aligned} P(\widehat{d} < d) &= P(\widehat{\theta}_{d-1} < \epsilon) \\ &= P(\theta_{d-1} - \widehat{\theta}_{d-1} > \theta_{d-1} - \epsilon) \\ &\leq P(|\theta_{d-1} - \widehat{\theta}_{d-1}| > \theta_{d-1} - \epsilon) \\ &\leq P(\|\widehat{K} - K\|_{\mathcal{S}} > \theta_{d-1} - \epsilon), \end{aligned} \tag{B.25}$$

where the final inequality follows from Lemma 4.2 in Bosq (2000). Now since $\theta_{d-1} > 0$ and $\epsilon \to 0$ as $n \to \infty$, it holds that $\theta_{d-1} - \epsilon > 0$ for large enough n. Thus, by (B.24) and an application of the Chebyshev inequality to (B.23), we have

$$P(\widehat{d} < d) \leq (\theta_{d-1} - \epsilon)^{-2} E\|\widehat{K} - K\|_{\mathcal{S}}^2 = O(n^{-1}). \tag{B.26}$$

From (B.24) and (B.25), it follows that

$$P(\widehat{d} \neq d) = P(\widehat{d} < d) + P(\widehat{d} > d) = O((\epsilon^2 n)^{-1}) \to 0.$$

This completes the proof. □

Acknowledgments. We are grateful to the Associate Editor and the two referees for their helpful comments and suggestions.

REFERENCES

Benko, M., Hardle, W. and Kneip, A. (2009). Common functional principal components. *Ann. Statist.* **37** 1–34. MR2488343

Besse, P. and Ramsay, J. O. (1986). Principal components analysis of sampled functions. *Psychometrika* **51** 285–311. MR0848110

Bosq, D. (2000). *Linear Processes in Function Spaces*. Springer, New York. MR1783138

Dauxois, J., Pousse, A. and Romain, Y. (1982). Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference. *J. Multivariate Anal.* **12** 136–154. MR0650934

Dunford, N. and Schwartz, J. T. (1988). *Linear Operators*. Wiley, New York.

Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis*. Springer, New York. MR2229687

Guillas, S. and Lai, M. J. (2010). Bivariate splines for spatial functional regression models. *J. Nonparametr. Stat.* **22** 477–497.

HALL, P. and VIAL, C. (2006). Assessing the finite dimensionality of functional data. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **68** 689–705. MR2301015

HYNDMAN, R. J. and ULLAH, M. S. (2007). Robust forecasting of mortality and fertility rates: A functional data approach. *Comput. Statist. Data Anal.* **51** 4942–4956. MR2364551

LEE, A. J. (1990). *U-Statistics*. Dekker, New York. MR1075417

LI, W. K. and MCLEOD, A. I. (1981). Distribution of the residual autocorrelations in multivariate ARMA time series models. *J. Roy. Statist. Soc. Ser. B* **43** 231–239. MR0626770

KNEIP, A. and UTIKAL, K. J. (2001). Inference for density families using functional principal component analysis (with discussion). *J. Amer. Statist. Assoc.* **96** 519–542. MR1946423

MAS, A. and MENNETEAU, L. (2003). Perturbation approach applied to the asymptotic study of random operators. In *High Dimensional Probability III* (J. Hoffmann-Jorgensen, M. B. Marcus and J. A. Wellner, eds.) 127–134. Birkhäuser, Boston. MR2033885

PAN, J. and YAO, Q. (2008). Modelling multiple time series via common factors. *Biometrika* **95** 365–379. MR2521589

PEÑA, D. and BOX, G. E. P. (1987). Identifying a simplifying structure in time series. *J. Amer. Statist. Assoc.* **82** 836–843. MR0909990

RAMSAY, J. O. and DALZELL, C. J. (1991). Some tools for functional data analysis (with discussion). *J. Roy. Statist. Soc. Ser. B* **53** 539–572. MR1125714

RAMSAY, J. O. and SILVERMAN, B. W. (2005). *Functional Data Analysis*. Springer, New York. MR2168993

RICE, J. A. and SILVERMAN, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *J. Roy. Statist. Soc. Ser. B* **53** 233–243. MR1094283

SEN, P. K. (1972). Limiting behavior of regular functionals of empirical distributions for stationary $*$-mixing processes. *Probab. Theory Related Fields* **25** 71–82. MR0329003

TIAO, G. C. and TSAY, R. S. (1989). Model specification in multivariate time series (with discussion). *J. Roy. Statist. Soc. Ser. B* **51** 157–213. MR1007452

N. BATHIA
Q. YAO
DEPARTMENT OF STATISTICS
LONDON SCHOOL OF ECONOMICS
LONDON, WC2A 2AE
UNITED KINGDOM
E-MAIL: n.bathia@lse.ac.uk
q.yao@lse.ac.uk

F. ZIEGELMANN
DEPARTMENT OF STATISTICS
FEDERAL UNIVERSITY
OF RIO GRANDE DO SUL
91509-900 PORTO ALEGRE
BRAZIL
E-MAIL: flavioaz@mat.ufrgs.br